

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 19:37:33 ; Search time 2187 Seconds

(without alignments)
11831.638 Million cell updates/sec

Title: US-09-613-486-14

Perfect score: 597
Sequence: 1 atgagatgcagtcgcgcagac.....gtagaagggagatgctaa 597

Scoring table: IDENTITY_NUC
Gapop 10.0 ; Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*
1: gb_da.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_pl.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_un.*
13: gb_un.*
14: gb_vl.*
15: em_da.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vl.*
30: em_hcg_hum.*
31: em_hcg_inv.*
32: em_hcg_other.*
33: em_hcg_mus.*
34: em_hcg_pln.*
35: em_hcg_rod.*
36: em_hcg_mam.*
37: em_hcg_vrt.*
38: em_sy.*
39: em_hlgo_hum.*
40: em_hlgo_mus.*
41: em_hlgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	597	100.0	597	6 AR138306	AR138306 Sequence
2	597	100.0	15000	4 AF039204	AF039204 Grapevine
3	597	100.0	15500	6 AR138299	AR138299 Sequence
4	593.8	99.5	8590	14 GLA4131	Y14131 Grapevine 1
5	374.6	62.7	16527	14 AF314061	AF314061 Grapevine
6	86.6	14.5	10545	14 BYU51931	US1931 Beet yellow
7	70.4	11.8	15468	14 AF190581	AF190581 Beet yell
8	65.6	11.0	15468	14 AF056575	AF056575 Beet yell
9	64.6	10.8	2724	14 BYVVC8P	M59452 Beet yellow
10	64	10.7	612	14 BYU71295	U71295 Beet yellow
11	63.8	10.7	6746	14 CLEBY3PH	X53462 Beet yellow
12	63.8	10.7	15480	14 BYVUA	X73476 Beet yellow
13	63	10.6	782	6 A41914	A41914 Sequence 1
14	63	10.6	5980	14 BYVMBPA	X73475 Beet yellow
15	37.8	6.3	2000	6 AX655393	AX655393 Sequence
16	37.8	6.3	151761	9 HSJ493D19	AL096868 Human DNA
17	37	6.2	145710	8 AP004765	AP004765 Oryza sat
18	36.8	6.2	1335	6 AX607511	AX607511 Sequence
19	36.8	6.2	29435	1 AE014256	AE014256 Streptococ
20	36.8	6.2	113050	6 AX602189	AX602189 Sequence
21	36.8	6.2	113050	1 SAG76851	AL76851 Streptococ
22	36.8	6.2	145254	2 AC114681	AC114681 Homo sapi
23	36.8	6.2	164155	9 AC000385	AC000385 Homo sapi
24	36.8	6.2	174171	1 AC068587	AC068587 Homo sapi
25	36.6	6.1	19103	1 AE007030	AE007030 Mycobacte
26	36.6	6.1	218193	2 AC145974	AC145974 Gallus ga
27	36.6	6.1	300050	1 BX248339	BX248339 Mycobacte
28	36.6	6.1	347436	1 BX842577	BX842577 Mycobacte
29	36.4	6.1	87942	9 HS31006	AL035593 Human DNA
30	36.4	6.1	168654	9 AC064878	AC064878 Homo sapi
31	35.8	6.1	625	8 AF462438	AF462438 Scleropha
32	35.8	6.0	142280	2 AC102011	AC102011 Mus muscu
33	35.8	6.0	198882	2 AC110577	AC110577 Mus muscu
34	35.8	6.0	206057	2 AC106183	AC106183 Rattus no
35	35.8	6.0	211062	2 AC126388	AC126388 Rattus no
36	35.6	6.0	237979	2 AC121719	AC121719 Rattus no
37	35.6	6.0	593	6 AR237648	AR237648 Sequence
38	35.6	6.0	155780	9 HS336012	AL513498 Homo sapi
39	35.6	6.0	175931	9 HS7386	AL513498 Homo sapi
40	35.4	5.9	179342	2 BX255923	BX255923 Human DNA
41	35.2	5.9	224166	2 AC110955	AC110955 Rattus no
42	35.2	5.9	264547	2 AC094432	AC094432 Rattus no
43	35	5.9	123210	2 AC136039	AC136039 Rattus no
44	35	5.9	141773	2 AC069348	AC069348 Homo sapi
45	35	5.9	160930	9 AC010093	AC010093 Homo sapi

ALIGNMENTS

RESULT 1
AR138306
LOCUS AR138306
DEFINITION Sequence 14 from patent US 6197948.
ACCESSION AR138306
VERSION AR138306.1 GI:14479815
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 597)
AUTHORS Zhu,H.-Y., Ling,K.-S. and Gonsalves,D.
TITLE Grapevine leafroll virus (type 2) proteins and their uses
JOURNAL Patent: US 6197948-A 14 06-MAR-2001;
FEATURES Location/Qualifiers

SOURCE 1. 597
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Query Match 100.0%; Score 597; DB 6; Length 597;
Best Local Similarity 100.0%; Pred. No. 4,7e-182;
Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGTTGATGTCCGACGACCACTTACCACTGTGTATACCGACGCTTACTACTA 60
Db 1 ATGAGTTGATGTCCGACGACCACTTACCACTGTGTATACCGACGCTTACTACTA 60
QY 61 AATGATGTCCGACCAAGCTTTATCTGCTGATGTAATGTTGTGTGCAAAAGG 120
Db 61 AATGATGTCCGACCAAGCTTTATCTGCTGATGTAATGTTGTGTGCAAAAGG 120
QY 121 GCTCTTACGAGGGTATAGAAAGTGTGTTGCTGCTACTCTTTACGCACTGCGGCAAGA 180
Db 121 GCTCTTACGAGGGTATAGAAAGTGTGTTGCTGCTACTCTTTACGCACTGCGGCAAGA 180
QY 181 ACCAGCTTCTTAAAGTTTACGCGGCAATTCAGACGTTATATTTTGAATAGTTTGGGA 240
Db 181 ACCAGCTTCTTAAAGTTTACGCGGCAATTCAGACGTTATATTTTGAATAGTTTGGGA 240
QY 241 GAGAGCAATGTGTATGTAACAGAGGTTGACCTTAAAGAGTACTGACGCGGTTGCGCCT 300
Db 241 GAGAGCAATGTGTATGTAACAGAGGTTGACCTTAAAGAGTACTGACGCGGTTGCGCCT 300
QY 301 CTCTAGTGTGCTTAACTTAACTTAAAGCTTGTGCTGCTACTTCTACTGAGGCTTACGTT 360
Db 301 CTCTAGTGTGCTTAACTTAACTTAAAGCTTGTGCTGCTACTTCTACTGAGGCTTACGTT 360
QY 361 GACTTTTGTATCGGCTATTAAGACCAATTAACCACTGACGCGGCGGCGGAATTTGGG 420
Db 361 GACTTTTGTATCGGCTATTAAGACCAATTAACCACTGACGCGGCGGCGGAATTTGGG 420
QY 421 ATTCAGCTGAAGATTCGCTACTTACGCTGACAGATTTTCTGGTACTTCCGGAAGCTCTCT 480
Db 421 ATTCAGCTGAAGATTCGCTACTTACGCTGACAGATTTTCTGGTACTTCCGGAAGCTCTCT 480
QY 481 GAATTAACGCAAGATGAGAAAGATGTTGCGGCTATGTAAGCTTAAACTGGAAGTGA 540
Db 481 GAATTAACGCAAGATGAGAAAGATGTTGCGGCTATGTAAGCTTAAACTGGAAGTGA 540
QY 541 GTGGAATATACACGAGTGAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 597
Db 541 GTGGAATATACACGAGTGAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 597

RESULT 2
AF039204 15000 bp RNA linear VR. 22-MAY-1998
LOCUS
DEFINITION
AF039204 Grapevine leafroll-associated virus 2 methyltransferase/helicase, putative transmembrane small hydrophobic protein, 65 kDa chaperone protein, 19 kDa protein, and 24 kDa protein genes, complete cds.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLES
JOURNAL
MEDLINE

AF039204.1 GI:3123909
Grapevine leafroll-associated virus 2
Grapevine leafroll-associated virus 2
Virus; ssRNA positive-strand viruses, no DNA stage;
Closteroviridae; Closterovirus.
1 (bases 1 to 15000)
Zhu, H.Y., Ling, K.S., Gosszyski, D.E., McFerson, J.R. and Gonsalves, D.
Nucleotide sequence and genome organization of grapevine leafroll-associated virus-2 are similar to beet yellow virus, the closterovirus type member
J Gen Virol. 79 (Pt 5), 1289-1298 (1998)
98264507

PUBMED 9603345
REFERENCE 2 (bases 1 to 15000)
AUTHORS Zhu, H.Y., Ling, K.S., and Gonsalves, D.
TITLE Direct Submission
JOURNAL Submitted (18-DEC-1997) Plant Pathology, Cornell University, New York State Agricultural Experiment Station, Geneva, NY 14456, USA
FEATURES
source
CDs
CDs
/organism="Grapevine leafroll-associated virus 2"
/mol_type="genomic RNA"
/db_xref="taxon:64003"
<1..7423
/function="replication and unknown functions"
/note="larger than 277 kDa; ORF1a, contains domains for two papain-like leader proteases, a methyltransferase and a helicase; identified by sequence comparison"
/product="methyltransferase/helicase polypeptide"
/protein_id="AAC4085.1"
/db_xref="GI:3123910"
/translation="ADVYAMLRVYCGKPELVMSRVYTPDRCYLAIRYLCAPYCP
PESRDVYALGMPVYARLACVKNPAGVAGIARGYTNRVYHCDYDAYYCYR
NLGRIIGGSDPDTSLTVISLGLGVPKNIAPFADIRYVPGGIGIYV
VTKREDHARVYERAAAGSVQPRQKIDERGCGVPSGSHLVGNLNEVRKVA
AGLIFRYGDMDFRSTQGHRLVYMRSSREVCLEIYSPKNIYRVLVPSGD
VAMSPFAAGRPFLVMTIRYRPNPQVLAHCRVACAFILRGPPKPDGAPPTA
KLNRNVSLEGRSLGMLYGYTRSGVPHCDYDAKFKDLRLMSAVIAGDCBEVY
PSDITPAMQKITEAVDRLYGGTSLKLKIDKIDKIDVOSLKDRIYKVPFY
MSEATQNSLTRFPQPELFSHSDSHPPAAASLLENEVLVRLCGNSVDIGCP
FALHSKTQRRVAVCDVLDKQACRVYRDLQYSVNRIGDDDKILEGRNIDIGYPL
GADHESAMMMVQYVDASILEICAMIKKSRITVMTGPELDRCEVMSLD
CEIEVDVADVYMGKSSCSYSHKSIKIDMTPEYTLIDGFLVEVYEMVAGNVP
KITSEVSPSICTLIRYRPNPQVLAHCRVACAFILRGPPKPDGAPPTA
EDYVYVNSANYSKTFEYVMSPIKSKSRVYISGKTIKIDVNLIDKTESRAVMA
GVRSLASVYLAUKLSHFGSDSFLVLESLKRLVLESLKRLVLESLKRLVLESLK
LDVSLDLDSTLSTLTPFAEKVLELEDELQCRABANERIKRLAAGSIAKALASEI
VVIDSKSPKQVNSSENADKREVCPGLRGSRNCGVGFLLFVVDASIRLFXVA
TDQRIKSYVFLDSVGFPLDYNVNLSEFLVSEGSYCMFAFLANNGLDSRVA
VCAKEVATSCANAVSAKVMITPAAVCAAMFVSCGSDGSEYKVIHRYOULF
DIFPESDSTYLPETLSAIGATVTLSESSSISLNAFLIOITKRGPELVYVNV
VTHGLSTAGTCVIRGVSQVSHLVNLTNVAYVQSFAGVPLVLYKRVSLITL
REDTSGTIGKISFPLSLFLKRLKDELSSITQGVFSDNRKVFATQBAR
TTWQVPAVAVDALSKAGKIYNNFTSRKLGKSSSDACSEYFSCDEBPGALK
GSSSYGFIILFESRIMGARRLVYKVECFGLPELSEKHEFRVFGKRTV
GVYDPLPTGIVETLSIECDQIELEIDDLKQKXDAISLDMNFESEDFLASIEEP
PPAGRGGSXKNIATILAEYANLFRIVYASCSKRPFLPAELSSALIEKFKVPR
KSOLVAIVREYTORFLRBRMALGINNEFVYKSPADILPALMKRVGSSFLASYRL
RGSYVCVSAARRRERFALVCTIGSLDPPVRYGACVCEVYASARPYRITIR
OKYVLSNPFCHBSVDSDSSAAGLKGASNMUTLPHLYRLASALISLOWESLKL
LSHNLFLCFALVDVAVNLIVLIGLSFFVQPLFSLPAAVLQPDREVESEKVLTA
FEFLKCSPPRALPKLFECVANSYVQVETKRVRLRCEVNLKLRGRLADRGTH
ROKAVPVPISNRVYVDGVERLSVMOQVGALETRLELIDEDSVIRKLNRRRDGN
DDEFTPPAHQMOEVTTPSCANAGLAEVAVEDAKSEKSKYKTEVVRGSGST
SEEVAFALSDDEAVELISVADERDQSPVYVISEYLNKINSFEPPIYVDKXDTG
GLNVAVERFYMOELPEIKSKICTYDQIRYINPRVSAPCEDQVLYRKGSGST
VQKGRVHIIDPHDHLTPGKISINRRRGAVLHNDNAFLASNLTPYKPSRS
EFTNNSVDILILEPFGGKTTILIDSLFVFKKGVSTVITLANKSVOYLAKVE
KEVNTIEQKRDKSPKSIYITDAVLMHRRGDADVLFIDEFMAHAGSLVACIF
TRCHKVMIFGDSROHYIERNELDKCIYGDIDRDVLOCRVYNI SVRCPMDVAGS
TVYGNLILVHEAGSETYARVNLVRLKPOEDPFPKSIHITVLRSHDSILVYVLA
RGDNTDCAIOKAAELVNRVFPPTSPGSSVINLVKRDVENSCKSSAPLVIYNF
LNEVNPRTAVIDPFDLSADPSTGPEFGCAASIVVRNINSSNITDHRKQV"

CDs
/function="replication"
/note="RdRp; 52 kDa; similar to RNA polymerases of other closteroviruses; presumably expressed via +1 ribosomal frameshift"
/product="RNA-dependent RNA polymerase"
/protein_id="AAC4086.1"
/db_xref="GI:3123911"
/translation="SVYRSCAIPRRKPSLOENLYSPFARNYVNFSTCDRNTSASMFGEA

MAMNLCRCPDLDAPSSLDVYISITRSGIEQMLEKRETSQIKALMKDVEPLEIDDE
ICRFKLMVADAKVADKSSCLTGHSAQNIIMFKRKSINAIFSPINFEVNRIMCCLKP
NIKFPMTNRDPAVSNMIGDDVYHIGVDPSKYDKSGADAFVAKVEVYKELGV
DEELAIAMGGERLSANTLDGSLFTIENKRSKGSANWTIGNSLVTGLISLYDVR
NBEALVIGDDSLIPSRISINADICTDMGPEKRMPSVPCSCVNMCGKPT
FPDPKIFVKACAVEDVSMPLFTFTSFQDLISDRFDEHLIQLEVALKTEVQ
TGATLALSVIHCLRSNLFSPSKLYPRVGMQVFTYSVKALKKSGCSLFDSEMTFG
QAMVWDEE
8865. .9035
/note="6 kDa; probably membrane-associated; similar to
small hydrophobic proteins of other closteroviruses"
/codon_start=1
/product="putative transmembrane small hydrophobic
protein"
/protein_id="AAC40857.1"
/db_xref="GI:3123912"
/translation="MNQVQLFECTFLINLAVFAVTPFILVFRVYKSPQKGRHAPV
PVVRGGGFSVTV"
9051. 10850
/note="p65; HSP70; similar to heat shock 70 proteins;
identified by sequence comparison"
/codon_start=1
/product="65 kDa chaperone protein"
/protein_id="AAC40858.1"
/db_xref="GI:3123913"
/translation="MNVFGLDPTFTSTVYVYKQGVSPKONNNAVITPTVYLRSDS
NMTFGFASLSMNLKVKGSFYRDLKRWGCGSSNLDLRLKPHISVRLKVGSG
LNEVTSIGNFGTVKSEBAHPLGLIALFIAVIAVISCAGAACCTGTGICVNPVYSVQ
RNFTDQCVLSGYQCVYMINPEPAALISQNSIGKSNAMAYDFGGGFDFDIISYR
NNTFVVRASGGDLNIGRDVRAFLTHLPSTLSEDLTIDLSNKESLKDADIVY
TLRVDGRKEDVAVNNKILTSVMLPYVNRITKLESTLSYKSNMESRACDLVLI
GSSSYLPGIADVLTKHOSVDRILRVSDPRAAVAGCALYSCGSGGGLIDCAHT
VALADSCHOICAPAGAPRPSGSMPLIARNNKSGQEVAFBEYIKCPKRNKIC
GAIIRFDIGVGTSTAPVPTFYMDSTISSGAVSPVVRPBGQVSLTPTNPSV
AAGSRVRELHILNNKVFGLILHRKARILLFTDEAIRVADSIDADIVKEKSY
AASALPDEDEVELLGRSVQKVLGRSLRLEIDP"
10777. 12432
/note="p63; putative heat shock protein 90 homolog"
/codon_start=1
/product="63 kDa protein"
/protein_id="AAC40859.1"
/db_xref="GI:3123914"
/translation="MNSYMSBLFKKYGADMKYLSRSHASHSEILKTLDIRLYG
GRVKSSEFSAALPNSPEOBLGLFLLSEEVGMSKLCITVEEAYDLINPAAVFTA
ETSPVYKSGCKYSMEDVNMWRILSNLDMLTEQCMSSGSELINDDGFR
VALTFKDRDPTADTGAANVECRGYDLVYAMSLFEORTKQSGNSLVEKXCEIAT
YLGSTDLPTADPRIPILNGILYDRCKEYVNPYSSYKRWDMRPRELAMYMLISDFV
VPMVYAPADVRLLELSEALELLEVPITSLIDSOYVGHILRYESTISDALDALE
DKREALKSNPRLSTAQAVGFCYIGERTKQSVVRPQVYKTPDSVSGFEINMK
DYKEFPLKQRELPNVSLRRQFGARHAFKIPKNGTISFPIISRLNVPREFWINTI
DYERANRSGLEBEILILNINISVDYRKCAERACNTLPSARFSSNHNKSNQSSROE
RRIKDLVLVKDLLEFQHRKAGWGSRSRDLGSRADHAKGSG"
12344. 13015
/note="p25; Cpd; coat protein duplicate"
/codon_start=1
/product="25 kDa diverged coat protein"
/protein_id="AAC40860.1"
/db_xref="GI:3123915"
/translation="MNSNTSVPVGGLAELETSGVLTTRKRAVDFKFNELKNENYSSV
DSRLKSTSEVLEKESKESFSELASTGBEHPYHILFPLICAKYLSIEKRYGSH
TYVVDKGTIVLDAMVPMKKSLTKRYKGVNGIARCCACEDLYLITVAIRERKRTK
AVGMKCLPVGKETLGNDFLSGTSKLSMDHRANYSIYAACNAVDRSAFTGGEKIVSLY
DLGRY"
13084. 13680
/note="p22"
/codon_start=1
/product="22 kDa coat protein"
/protein_id="AAC40861.1"
/db_xref="GI:3123916"
/translation="MELMSDSNLSNVLITDASLNGVDKLLSAEYKMLVQKAPNE
GIVVFGLLYALAAFTSPKQVADSDIFNSFGBERVVVTEGDKLVLDGCAPLT
RFTNKLRTFGRTTEAVVDFCIAYKHKILPOLNAAELGIPADBSYLAADFLCTCPKLS

ELQSRKFAFAMTALKTGGVNTPVSNLRQLGRREV"
13680. 14165
/note="p19"
/codon_start=1
/product="19 kDa protein"
/protein_id="AAC40862.1"
/db_xref="GI:3123917"
/translation="MEDYEKSESLILRLTNLTMLVKSASVELPKLLIGYLRV
SGRGEVTCNRBELTRDPEGNHHTVLSRIQYDSSEAFEEFNNSDVVFPLETGSV
FMPFLRSETGRVAVRHILRTEFPAANFFPGSHCGMEYCLKQVLTETESLIDSCERN
R"
14167. 14784
/note="p24"
/codon_start=1
/product="24 kDa protein"
/protein_id="AAC40863.1"
/db_xref="GI:3123918"
/translation="MRVIVPYEAEEDILKRSITMLRVISDGLSTKECIKAFSTTRD
LHCAKSYQNGVDGILQKNCABERKLDITVSNIRLAQPLVREKVAHPRCKDRELY
AFIRKIVELTGVRKREKRSILTKVILNKRSLMARVMSFRAMVEMLEKFS
PVKIFRDLIDVETINELCAEDVDVKNVENEDENHDLQDEC"
14785. 15000
/note="similar to 3' UTRs of other monopartite
closteroviruses"
3'UTR
ORIGIN
Query Match 100.0%; Score 597; DB 14; Length 15000;
Best Local Similarity 100.0%; Pred. No. 7,7e-182;
Matches 597; Conservative 0; Mismatches 0; Gaps 0;
1 ATGGAGTGAAGTCCGACAGACCTTAAGCACTGGTGAATACCGACGCTTAGTCTA 60
13084 ATGAGGTGAAGTCCGACAGACCTTAAGCACTGGTGAATACCGACGCTTAGTCTA 13143
61 AATGGTGCACAAAGACCTTTATCTGCTGAAGTGAATAAATGTTGGTACAAAAGG 120
13144 AATGGTGCACAAAGACCTTTATCTGCTGAAGTGAATAAATGTTGGTACAAAAGG 13203
121 GGTCTTACGAGGGATATGAAGGTGTGTGGTGTACTCTCTTACGACCTCGCGGACAGA 180
13204 GTCTCTTACGAGGGATATGAAGGTGTGTGGTGTACTCTCTTACGACCTCGCGGACAGA 13263
181 ACCAGCTCTCTTACGAGGGATATGAAGGTGTGTGGTGTACTCTCTTACGACCTCGCGGACAGA 240
13284 ACCAGCTCTCTTACGAGGGATATGAAGGTGTGTGGTGTACTCTCTTACGACCTCGCGGACAGA 13323
241 GAGAGGAATGTGTATGTAACAGAGGTGACCTTAAGAGGACTCGACGGGTGGGCGCT 300
13324 GAGAGGAATGTGTATGTAACAGAGGTGACCTTAAGAGGACTCGACGGGTGGGCGCT 13383
301 CTCACAGGTTCACCTAATAAATTGAACGTTGGTGTGTACTCTTACGAGGCTTACGTT 360
13384 CTCACAGGTTCACCTAATAAATTGAACGTTGGTGTGTACTCTTACGAGGCTTACGTT 13443
361 GACCTTTTATCGCGGTATAGCACAAAATTACCCCACTCAACGCGCGGGGGAATTGGG 420
13444 GACCTTTTATCGCGGTATAGCACAAAATTACCCCACTCAACGCGCGGGGGAATTGGG 13503
421 ATTCAGCTGAAGATTGCTACTAGTCTAGAGATTTTGTGGTACTTCCCGAAGCTCTCT 480
13504 ATTCAGCTGAAGATTGCTACTAGTCTAGAGATTTTGTGGTACTTCCCGAAGCTCTCT 13563
481 GAATTACGCAAGTAGAGAGATGTTGGCGAGTATGTAAGCTCTTAAAAAATCGAAGGTGGA 540
13564 GAATTACGCAAGTAGAGAGATGTTGGCGAGTATGTAAGCTCTTAAAAAATCGAAGGTGGA 13623
541 GTGTAAATACCAAGTAGAGATGTTGGCGAGTATGTAAGCTCTTAAAAAATCGAAGGTGGA 597
13624 GTGTAAATACCAAGTAGAGATGTTGGCGAGTATGTAAGCTCTTAAAAAATCGAAGGTGGA 13680
RESULT 3
AR138299

LOCUS AR138299 15500 bp DNA linear PAT 16-JUN-2001
 DEFINITION Sequence 1 from patent US 6197948.
 ACCESSION AR138299
 VERSION AR138299.1 GI:14479808
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 15500)
 AUTHORS Zhu H.-Y., Ling K.-S. and Gonsalves D.
 TITLE Grapevine leafroll virus (type 2) proteins and their uses
 JOURNAL Patent: US 6197948-A 1 06-MAR-2001;
 FEATURES
 source 1. 15500
 /organism="Unknown"
 /mol_type="unassigned DNA"

ORIGIN

Query Match 100.0%; Score 597; DB 6; Length 15500;
 Best Local Similarity 100.0%; Pred. No. 7,7e-182;
 Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGAGTTGATGTCGACAGCAACCTTACGACCTGGTATTAACGACGCTTACTCTA 60
 Db 13584 ATGGAGTTGATGTCGACAGCAACCTTACGACCTGGTATTAACGACGCTTACTCTA 13643

QY 61 AATGGTGTGACAGAGAGCTTTATCTGCTGAGAGTTGAAAAAATGTTGGTCAGAAAGG 120
 Db 13644 AATGGTGTGACAGAGAGCTTTATCTGCTGAGAGTTGAAAAAATGTTGGTCAGAAAGG 13703

QY 121 GCTCTTAACGAGAGGTATAGAGTGTGTGCTTACTCTTACGACCTCGCGGCAAG 180
 Db 13704 GCTCTTAACGAGAGGTATAGAGTGTGTGCTTACTCTTACGACCTCGCGGCAAG 13763

QY 181 ACCACGTCCTCTTAAGGTTACGCGCGAGATTACAGCTTATATTTCAAATAGTTTGG 240
 Db 13764 ACCACGTCCTCTTAAGGTTACGCGCGAGATTACAGCTTATATTTCAAATAGTTTGG 13823

QY 241 GAGGAGATGTGTAGTAACAGAGGTGACCTTAAGAGGTACTCCAGCGGCTGCGCCT 300
 Db 13824 GAGGAGATGTGTAGTAACAGAGGTGACCTTAAGAGGTACTCCAGCGGCTGCGCCT 13883

QY 301 CTCACCTAGTTCACTAATAAAGTTAGAACGTTGCTGCTTACTTCACTAGGCTTACGTT 360
 Db 13884 CTCACCTAGTTCACTAATAAAGTTAGAACGTTGCTGCTTACTTCACTAGGCTTACGTT 13943

QY 361 GACTTTGTATCGGTATTAAGCAAAATTAACCCCACTCAACGCGCGGCGAAATGGGG 420
 Db 13944 GACTTTGTATCGGTATTAAGCAAAATTAACCCCACTCAACGCGCGGCGAAATGGGG 14003

QY 421 ATTCACAGTGAAGATTGCTAGTACGAGATTTTGTGGGTAATTGCCGGAAGCTCTCT 480
 Db 14004 ATTCACAGTGAAGATTGCTAGTACGAGATTTTGTGGGTAATTGCCGGAAGCTCTCT 14063

QY 481 GAATTAACAGCAAGTAGAAGAAATGTGGCGAGTATGACGCTCTAATAACTGAAGGGA 540
 Db 14064 GAATTAACAGCAAGTAGAAGAAATGTGGCGAGTATGACGCTCTAATAACTGAAGGGA 14123

QY 541 GTGGTAATACACAGAGTAGCAATCTGCTAGCTAGTAGAAGGAGTTATGTAA 597
 Db 14124 GTGGTAATACACAGAGTAGCAATCTGCTAGCTAGTAGAAGGAGTTATGTAA 14180

RESULT 4
 GLAV4131 8590 bp RNA linear VRL_08-JAN-1998
 LOCUS Grapevine leafroll-associated virus 2 genes encoding RNA polymerase
 DEFINITION and coat protein, hsp70, hsp90 gene and ORF2, ORF7 and ORF8.
 ACCESSION Y14131
 VERSION Y14131.1 GI:2765391
 KEYWORDS coat protein, heat shock protein 70-like; ORF1; ORF2; ORF3; ORF4;
 ORF5; ORF6; ORF7; ORF8; RNA polymerase.
 SOURCE Grapevine leafroll-associated virus 2

ORGANISM Grapevine leafroll-associated virus 2
 REFERENCE 1
 AUTHORS Abou-Ghanem N.
 TITLE The nucleotide sequence of the 3' terminal region of grapevine leafroll associated closterovirus 2
 JOURNAL Unpublished
 REFERENCE 2
 AUTHORS Abou-Ghanem N.
 TITLE Direct Submission
 JOURNAL Submitted (01-JUN-1997) N. Abou-Ghanem, Universitat degli Studi di Bari, Dipartimento di Protezione delle Pianta e Centro studio virus e virosi delle colture mediterranee, via Amendola 165/A, I 70126 Bari, ITALY
 REMARK 3 (bases 1 to 8590)
 REFERENCE 3
 AUTHORS Abou-Ghanem N.
 TITLE Direct Submission
 JOURNAL Submitted (07-JAN-1998) N. Abou-Ghanem, Universitat degli Studi di Bari, Dipartimento di Protezione delle Pianta e Centro studio virus e virosi delle colture mediterranee, via Amendola 165/A, I 70126 Bari, ITALY
 COMMENT On Jan 10, 1998 this sequence version replaced gi:2369864.
 FEATURES
 source 1. 8590
 /organism="Grapevine leafroll-associated virus 2"
 /mol_type="genomic RNA"
 /db_xref="taxon:64003"
 /lab_host="N. benthamiana"
 1165..2427
 /note="ORF1"
 /codon_start=1
 /product="RNA polymerase"
 /protein_id="CA474561.1"
 /db_xref="GI:2369865"
 /db_xref="GOA:O39852"
 /db_xref="SPTREMBL:O39852"
 /translation="MPGAMAMNCLRCDFLDAPSLRNVDYISITRSGIPEQLKRP
 SRKAMKDVESPLEIDELCRFLKAMKRDVAKVLDSSCTKRSAGNINERHSIA
 IESPIENVARIRINCKLRPKIKFTFTNDRDPAFVAVNAGDDPVYIGVDESKYK
 SODAFVRAFEVMTRELVGDELLAIVMGSRSLKATNTDQGLFTIENRKSASNT
 WIGNSLVTLGLSLYVDVNEALYIGDLSLRSRLISYADICTDGFETKREMS
 PSVPYCFKFMVCGHKTFVLDPYKLFVKLGAVKEDVMSDSFETFSRDLTSDN
 DERLIQKLAELVALKEYQTGNTLALSVHCLRSNPLSFESKLYPRVKGMQVFTSVK
 KALLKSGSLPDSFMTFGCAVMWVDE"
 2491..2661
 /note="ORF2"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="CA474562.1"
 /db_xref="GI:2369866"
 /db_xref="GOA:O39853"
 /db_xref="SPTREMBL:O39853"
 /translation="MNQVLQFECLFLNLMAVFAVTFIFILLVPRVKSFRQGREAPV
 PVRRGGSTIV"
 2677..4476
 /gene="hsp70"
 /gene="hsp70"
 2677..4476
 /gene="hsp70"
 /codon_start=1
 /product="heat shock protein 70-like"
 /protein_id="CA474563.1"
 /db_xref="GI:2369867"
 /db_xref="GOA:O39854"
 /db_xref="SPTREMBL:O39854"
 /translation="WVVGDLDFGFTFSVVCYKDGRRVSEFKONNSAVIPTYLVLPSSD
 NMTPGYAESLMSNLKVGSEFYDLKMGVCDSSNLDYADRLKPHYSVLVYIGSG
 LNEITYIGNFGVYSEAHLPGLTALFTKAVISGEGAFACGTCGTCVPAVNDYVQ
 NPTDQCVISGYOCVWIMNIPSAALSAQSVKRSANLAVYDGGGTGFTVSIISYR
 NNTFTVYRSGDINLIGSDVDRAPLTHFLFSLTSEPDITLYSNLKSGLSTDTDEIY
 TURGVGKREDVAVKNNLIVMLFTVAKTILISTLSYKASLKNBSALVKCDLVLT
 GSSSYLPGLADVLTRHGSVDRLRVSDPRAAVAVGCAVYSKLSGSGGLTIDCAHT

VAIADRSCHQIICAPAGAPIPPSGSNPLYLAAVNKNKSGREIAVEFGEYVKCPKKNKIC
GNIIRFIDIGVGDYAPVTFYMDPISISVGVASVVRPBEKQVSLGTGPAVNSV
AAGSRSVRLHISNNKVFGLHRLKRAHRLFTDEAIRYADSIADIADVLKESY
AASAPPEDEVELLKGKSYQVLRGSRLEIPL"

gene

CDS

/gene="hsp90"
/codon_start=1
/product="heat shock protein 90-like"
/protein_id="CA74564.1"
/db_xref="GI:2369868"
/db_xref="SPTREMBL:O39855"

/translation="MSNYGMSLFFKFTYEADMKYLSHSTAHSSEIKTLPDILYG
GRVKKSEFESALPNSFEOLGLFISREVEGMSKCGITVEAAVDLNPAPVPTA
ETGSPVKGQGYQSMEDVNCVRLNSLVNDKMLAQCMSLNSCGELVNDGGRF
VATLPDRTADDTGAANVECRVGYLVYVAMSLEORTQKSGNLSLEYKCEYIRT
YLSGTDLPFAPDRIPLTGLIHDPCKEYNI FYSYKRVNDRFELAYMPLISDF
VPCWMAAPDPRLLFELISAELILEVPTSLIDSOVVGHLIRVYESTSDAIDALE
DKLEIKSNPRLSTAOJMWGPFYCGEFTPAOSVYVORPGVYTPSPGSEFINMK
DYEKFPDKORELPNYSILKROKRAHAFKIKNGNISFPISRLINVPREFWTINI
DIFRANRSGLTSEEILINNISVDVRKLCARACSTPSARFSPNHNKSNQSROE
RRKQDLYVVKLTLYEFOHKKRAGWGSRTLDGSRADHAKGSG"

CDS

/note="ORF5
duplicate gene"
/codon_start=1
/product="coat protein"
/protein_id="CA74565.1"
/db_xref="GI:2369869"
/db_xref="GOA:O39856"

/db_xref="SPTREMBL:O39856"
/translation="MSSNTSVPGVGLAELETSGVLTTRKEAVDKFNEIKENYSSV
DSRLSDSVKVTLEKSKESFSLSTDEHVVHILFELICAKISSEKRYGSKX
TYVVDKRTYVLDAYFNMKSLTKRYKRVNGIRAPCCCEPLIYTAIPINSEKXIK
AVGMKGLPVGEKYLGLADFLSGTSLKMSDHDRAVSTVAANAVDRSAFTGGERKIVSLY
DLGRY"

CDS

/note="ORF6"
/codon_start=1
/product="coat protein"
/protein_id="CA74566.1"
/db_xref="GI:2369870"
/db_xref="GOA:O39857"

/db_xref="SPTREMBL:O39857"
/translation="MELMSDSNLNLTIDASSLNGVDKKLSAEVVKMLVOKGAPNE
GIEVFGLLVLAARTSPKVOADSDVTFNSPSEERVVVTEGDKKVDGCAPLT
RFTNKLRTGRTTEAYVDCIAYKKKLPOLMAAAELIGTPADSVLADFLGCTCKLS
ELQGRKMPASMAKLTBGSVYNTPVSNLRQGRBWM"

CDS

/note="ORF7"
/codon_start=1
/product="hypothetical protein"
/protein_id="CA74567.1"
/db_xref="GI:2369871"

/db_xref="SPTREMBL:O39858"
/translation="MEDYREKBSLILNTNNTMLLVKSNASVELPULICGLIYR
SGGEVTCNRELTDFGHNHTVIRSRILGYDSSAAEEENNSDCCVAFLELEGSV
FWFELSETKRAVRLRTFFEANNFPSHCGTWEYCIKQVLTETESLIDSCBERN
R"

CDS

/note="ORF8"
/codon_start=1
/product="hypothetical protein"
/protein_id="CA74568.1"
/db_xref="GI:2369872"

/db_xref="SPTREMBL:O39859"
/translation="MEVIVSPYAEADILKRTSDMLRNIDSGVLSKCEIKAPSTTRD
LHAKSYOMGVDTGLYORNCAEKRLIDVENSIRLAQPLVREKAAVHPCQPEPELV
AATRKCYELTSGREAYKREKRSITKIVLKNKSLTMAFWSPAPAKMAEVLKFS
PVATFRDLADVELTINELCAEDDVHDKVNEDENDHDLQDEC"

ORIGIN

Query Match 99.5%; Score 593.8; DB 14; Length 8590;
Best Local Similarity 99.7%; Pred. No. 7,7e-181;
Matches 595; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	1	ATGAGAGTGAATGTCGACAGCAACCTTAAAGCACTGCTGTAACCCGAGCCTAGCTA	60
DB	6710	ATGAGAGTGAATGTCGACAGCAACCTTAAAGCACTGCTGTAACCCGAGCCTAGCTA	6769
QY	61	AATGGTGTGACAGAAAGCTTTTATCTGCTGAAGTTGAAAAATGTTGGTGCAGAAAAGG	120
DB	6770	AATGGTGTGACAGAAAGCTTTTATCTGCTGAAGTTGAAAAATGTTGGTGCAGAAAAGG	6829
QY	121	GCTCCTAACGAGGGGTATAGAAGTGTGTGCTATCCTTTACAGCACTCGGCGGACAGA	180
DB	6830	GCTCCTAACGAGGGGTATAGAAGTGTGTGCTATCCTTTACAGCACTCGGCGGACAGA	6889
QY	181	ACCAAGTCTCCTTAAGGTTACGCGGACAGATTACAGCGTTATATTTCAATAGTTTGGGA	240
DB	6890	ACCAAGTCTCCTTAAGGTTACGCGGACAGATTACAGCGTTATATTTCAATAGTTTGGGA	6949
QY	241	GAGAGGAATGTGTGATGAACAGAGGAGTGAAGAGTATCAAGCGGTGTGGCCT	300
DB	6950	GAGAGGAATGTGTGATGAACAGAGGAGTGAAGAGTATCAAGCGGTGTGGCCT	7009
QY	301	CTCACTAGGTTCACTAATTAACCTTACAGCGTGTGTGCTGCTTCACTGAGGCTTACGTT	360
DB	7010	CTCACTAGGTTCACTAATTAACCTTACAGCGTGTGTGCTGCTTCACTGAGGCTTACGTT	7069
QY	361	GACCTTGTATCGCGTATAGACACAAATTAACCCCACTCAACCGCGGCGGAAATTTGGG	420
DB	7070	GACCTTGTATCGCGTATAGACACAAATTAACCCCACTCAACCGCGGCGGAAATTTGGG	7129
QY	421	ATTCAGGTGAAGATTGTAAGTCTGCTGAGATTCTTGAGGTACTTCCCGAAGCTCTCT	480
DB	7130	ATTCAGGTGAAGATTGTAAGTCTGCTGAGATTCTTGAGGTACTTCCCGAAGCTCTCT	7189
QY	481	GAATTACAGCAAGATAGAAAGTGTTCGAGATATGTAACCTCTTAAACTGAAAGTGA	540
DB	7190	GAATTACAGCAAGATAGAAAGTGTTCGAGATATGTAACCTCTTAAACTGAAAGTGA	7249
QY	541	GTGGTAATTAACAGCAGTGAATCTGCTGCTGCTAGTGAAGGGAAGTATGTAA	597
DB	7250	GTGGTAATTAACAGCAGTGAATCTGCTGCTGCTAGTGAAGGGAAGTATGTAA	7306

RESULT 5
AF314061
LOCUS
DEFINITION
Grapevine rootstock stem lesion associated virus
Grapevine rootstock stem lesion associated virus
polymerase, p6, HSP70-like protein, HSP90-like protein, coat
protein duplicate, coat protein, p19, and p24 genes, complete cds.
AF314061
AF314061.1 GI:24636914

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Grapevine rootstock stem lesion associated virus
Grapevine rootstock stem lesion associated virus
Viruses; ssRNA positive-strand viruses, no DNA stage;
Closteroviridae; Closterovirus.
REFERENCE
1 (bases 1 to 16527)
Zhang, Y. and Rowhani, A.
Nucleotide sequence of grapevine rootstock stem lesion associated
virus
Unpublished
2 (bases 1 to 16527)
Zhang, Y. and Rowhani, A.
Direct Submission
Submitted (16-OCT-2000) Plant Pathology, University of California,
One Shields Ave., Davis, CA 95616, USA
FEATURES
source
1.16527
/organism="Grapevine rootstock stem lesion associated
virus"

```

CDS
/mol_type="genomic RNA"
/db_xref="taxon:167634"
/size="8902
/notes="327 kDa"
/codon_start=1
/product="methylin transferase/helicase polypeptide"
/db_xref="GI:24636915"
/translation="MSSLAISLPSAVAOISFQOPATYAKSFYMTSLPSLOTCTPQC
EFSLSFICLASFOKIKTFLFSLVRELAFLSELSTGCEPIRLGVEHPLFFVFSF
QCGPLSYCLGCVNGIFAVNLISFPLDVLASLMDVDVVTPTVIRISKRKAKLAKAF
SIAAARIRSRKXGKNEVISHQPLPACSLFSONKGTLELIIGRVGVATYAT
KAPVRRGGGGRVPLIPLVAFSGVYCAPSCGALACGKSLASVANDSAE
KLSAPRRKIOONENSRVAKADVTYASVDMAEGEATPPRIFGSGSEPPPISS
SGMARSLREPAIVTCGASAPAKETPRYSYKADPSAPAPRPAALPRVAKREA
IIVGVVHVLNKRISGVLPFRVGDMRSPFSSTHGCKIVANVTINSMNCTKV
NGEFPVALPCSDVYAMLRVYSGKPEPLVMSRECYPDGLCYAMHRYLAFECQF
RESYDPLGSPVVALKALVIRFGDEALNIGVGYVSRARFACDYDVARSLMRL
AGYVGDVHISTLASITVYLKPEPADRPKISFGSCCELRYESVNDAPRLSE
KULRVPRNGVRESADVGSNTVRKADRIDKRFSLLAGSVNKKVKAIAAGVLYR
VGGMDPHNSFLTOAGYHILYRKTSISVCELPDTPGRLPRDVI PCSDVAMMRF
AAGRPVLVLTDRKTKNGCYTAAHCHYASAPILKGRHVPDPTGAPTRAKLRMY
SVIGDRSLNLTGYSFTSRGI FHCIDYDAVKDRMSALVAGKEMGVASDITPA
MKQTEAVYDRLYGSNDLKLKSIKEDLIDFNDOVSLKQKRPVIRPEYSEVQIN
SLTRFYQPELKFESHSDHPAAASRLNEETLVRLCGHSVDIGGCPFLNKSX
ORRVHCRPVLGDKAQRVYRELQYSVRSVDEKLEGPNDVCHPLGCHEHS
STMMVQVYASALYELCSAMIKKKARITLYLTMTPGPDLGRCYMESLDCDIEYD
HAYVMYKFGSGSCYSHKSLIKDITMTPTVLTGGLPSVEMYRIMGVANVPKTKSEV
CPNISCKLRRPRANSDDVYKLPREDKRRMCLGSDITIIDSKFVSREPTVYCN
CSAANKSTPEWMSFVSKSRVITSGKTIHKDYNLDLKVESFAVAMLASGVRSFLA
SEYLANKLHSFGDGFSEIATSPVLRREINMTINFECLLVKRAVATLDSVSLD
LDAETESITDFAECRVAVDLELGLANANEKIRLADSMANALIVDIDIAK
PVAHAGSEVESEKHAQPLRGSGKAGVLEFLDPTLSASRLFGSCVDAQOQIK
SVYHILTAVSFLDONTENLSPIMRVISKGYSCMFYANRGLANRTOULPDITFED
AUCANYSVSKAKYTPYPSAAGCMVSCGSCGSCGSCGSCGSCGSCGSCGSCGSCG
SSVPLTBLSSAICGAVITLPSGTSYSLNAPLQITKFSLELIRANVRYTHSLST
TADGVRSVFSQIISHLVNGTGNLAVQIAPLGVPLLVKCVGLITFMSDITSG
PIKGVSEFPTINILKYRGLQVDELKAVQVFPDKNKLPDVTQEARLRSVIMQIP
TAAVDGLKCAKASVSGIVGKTNVGEKRGDGDHEEYFCSDBESTPDLRGSGSG
FSILFSPRLVRRARLIGRIKRFCSKIPESISLKHLETLFLGKSGSDVGYEFL
SEVDYALSSIEEDELINLIGDFFRMDADSLDMNYPDSEDFLASVNGACAGL
GCGRRVGMALINVTYNSLIMAAASKNSTPTSLIFRGLVDRIANITFNSSEVVE
LVKRYVNFILSKRKGILGNHRYNSFGELVPLLNKLSRTIYRAVMSKCKVY
CLASERKEFALCVLTGLPYPFARLVGAVCDKLSLRRAVDRIYFLKSCDA
MNGRLCLFHFDDNGDSTAGLKGGSGQTFWGLLRLVRLALSGLYAGCQVSHDL
FLFCVSLVEEYMAKILIGLSEFSPRISALAFALQPERFVEYSGKSTTLFELIK
ISPRVPLIKGFEMVAGSPILRGVRRFLIAVKILRRKGLRTNGSPQORAPV
VIPSRYTNGDCERLSEKMOGVEALREELILELDAVVEKLNKRNRREADESER
PMHSTHTDITSFCSKANSVGTALERAVLVEADVAKSEKISPMARDIVAKGTVDSEITN
TLSDDEALIEVPVSEBDDPKTYRVESEYLNRLNLSFEPKPIYVNDNETGGLTNAY
REFYMOELAFELHSLCAITDOLIVNERSMAPCNEDQIVRNGSLITQGVQY
RSNICKFHDHFLDGKISVNRRRSGNVLYHDLSPLSANLPLAGPEPERSVFNIS
SVLILYEPGGGKITTILIDSPKVFKEGVSEFMTLANKSSQVELLKVEEDINIS
ECQKDKRTPPKSITITIDAVIMHRCADILFIDCFMVAHAGSVLACKEFCHKV
MIFGDSQIHYIENELDKCYGLDFPDVLOGGVNYSYRCPMDVCAULSVYNNM
IATVGSSEKSSMRINEINSVDLVDPMGSTYLCMOSKLEISHFIRKSGSKLVN
LTVNAGECTYARVNLVRLKQDEPFPKSRHITVALASRPTDSITVTVLAARGDATC
DAIOKABELVAKRVPFPTSGGSYINIDVAKDEDNRCGASAPUSVINDPFLDVNP
GTVITDFGDLSDSTGSPFEGGASGIVARNISSNLTDDKORV"
/size="8901.10280
/notes="53 kDa; translated via ribosomal frameshift"
/codon_start=1
/product="RNA-dependent RNA polymerase"
/protein_id="AAN63467.1"
/db_xref="GI:24636916"
/translation="SAVRSQAIIPRRKSLQENLYSPEARNYNSTCDRTFSAMFSEA
MAANCIKRCFDLETFSSLRDVISITRSGIEOMLEKRTPSQIKALMNDISPEIDE
ICREFLWYKRAKYLDSSCLTKSPANIKFHKSNALFSPIFNEVKRIRKINSCKP
NIFKFTETNRDPAVSNNMGGDDVYHIGVDSPSKYDQDAFVSFEFMEKELV
DELLAIWMCGERLSTANTLGO.SFTINORSGASNTWINGSLVTLGLISLYYDR
HFPAIYSGDGLIFSRKINRYADICADMGPEPKSPSVYFCGKPFVMVGHGTF
FVDPDYKLVKLVKEDVSNDF.PETFTSGKQVTVYVKALEFSGGSLVESYVTPGS
SGNTALASVTHCLRSNPLSKLYPRVKGQVYFYSVKALEFSGGSLVESYVTPGS

```

```

CDS
HAWVWMDDE"
10373..10543
/size="6 kDa"
/codon_start=1
/product="p6"
/protein_id="AAN63468.1"
/db_xref="GI:24636917"
/translation="MDQVLOFECVYLLNQAVFATFPVILLIRIVKSPRRNHEPTI
SAVSGQSPSTV"
/size="10552.12351
/notes="65 kDa"
/codon_start=1
/product="HSP70-like protein"
/protein_id="AAN63469.1"
/db_xref="GI:24636918"
/translation="MVYFGLDFGTPTSTVCYVYHDKGYKSPKONNSAYIPVLYLFSPT
NHMFVGAESLIMFNKVRGSEFFDLKRWGCDSSNFDYDRLKRYAARPKYIGTG
LNDVITIGNYGVTRSEVHLPEGLIASFIKAMVSCAENAFSCGTGVICSPANDSVQ
RNFIDQCVSLSGYOCVYVWMEPSSAALASCNLIKMSANLAVDPFGGIDVSLISYR
NNTFVYRASGGDLNIGSDVDRPTLTLFSLTSLSPDLSDISNLKESLKTDAEIVY
TLKSGDDKEDVRYNKSILITVWLPYNRITLILEATLKSAYKMGGLRSYCDLVLI
GSSVYLGLELIDILSKRIRIDRIKADPRAAVNCALYSKCSGSGGLLIDYCAAT
IATKMSYCDQITCKPAGAPLPFGSTTPLYLPKARNQORQIAYEFGYVYCPRRKIC
GSNTRFLDITGVTSYSFTFYLDVTISSGALSFVYRGSGEVSITGTPTAYNPSYV
VLGRSVREELHLSNKHVGLLHRRADRRLLESEGEAVRYETVEADVLEKFSY
NASSLPDEDEVEPLMGKSVQYVLRGRLIEEIP"
/size="12278.113882
/notes="62 kDa"
/codon_start=1
/product="HSP90-like protein"
/protein_id="AAN63470.1"
/db_xref="GI:24636919"
/translation="MNSHNSWGLSPFKYKYGENDWERYLSKSIVAHTANIKTLDPVQLYG
GRYKYSSEFATPSFEQELGLMLSEBDGWMKMGITVEEMVQITNDEAPIFA
EKCPDYKTEGQKTMEDVNMGRNSMLDINDKLIACQMSLSCGGLINDPDKEFA
VALTRKNEIYVNSNNTIDCGVDGYVYAMSLFNGCSQNSQAGNISLYEKYCAIYR
YIENTDYIFTEPDPILPILITGLIYPCCEYNIIFYSTYRNNDNPPFPLTNMPLISDI
VFQWKPXTVDRLPLFLISAEULTVEPLLSVDSQVVIQVLRVVESTADPAIDALE
EKDLAMKSSPHSLTAQJLWGFYCYGEFTAGRAVQIRGVKTPNSVGGFETNMK
NVEEFDDITQIEVYVNSLRQFNQARHAEAKVKKQNISEPKISRLNIPREFYLVN
DYFRANRAGSEEEVILINNIYDVRLCAERACSTLPSAKRFSKHSKSVPSLRQ
KYRPLPILANSLSEYFHHGRGRV"
/size="13845.14516
/notes="25 kDa"
/codon_start=1
/product="coat protein duplicate"
/protein_id="AAN63471.1"
/db_xref="GI:24636920"
/translation="MSSGTGJAVGSESEALIEGTVLSTRKEAVDKFENELNENYASY
DSSRLSDQEVLEKSKSEPSKSLASDEHPYHIIFFLIRSAIISGDSKYVGS
KYVDEKTYTQDVQVWFMIMIGLTKOYRVAAGLAFCCACEDYLITVAPAMSEFKK
AIGKGLVGVGEYIADLPLSGTSKLMSDHDAAVSITVAKNAVDSAPFTGGRKIVSLY
DLGR"
/size="14604.15200
/notes="22 kDa"
/codon_start=1
/product="coat protein"
/protein_id="AAN63472.1"
/db_xref="GI:24636921"
/translation="MELMSDNLISGLVITTDASSINGVDKLSAEVITKMLYOKAPSO
GIEYFGLILVALAARTTSPVQADSDIIFQNTYGDVTVVATGDLKVEGCAPLN
RFTNRLKTFGTTEAYVDPCVAYKHPQUNAAAEVILPDESDYLADFLGACPKS
ELQGRKRFASNYALKTEGVYNTFVSNLRLQGRREV"
/size="15200.15685
/notes="19 kDa"
/codon_start=1
/product="p9"
/protein_id="AAN63473.1"
/db_xref="GI:24636922"
/translation="MEICQNPDEALVLTNTQNTLLVYKSDDEVNLEKLLICGYLRY
SHRGDVTCNEDVYKDEGAHHVIRERTVRYVESAVKEYNADCVKFFLETDGV
RWFFIQSDIKRRAAHRLTFPEANPPFGSGHGMVEYCLKVLLITEVESVPSCEBRN
R"

```

CDS
15687..16304
/note="24 kDa"
/codon_start=1
/product="p24"
/protein_id="AA063474.1"
/db_xref="GI:24636923"
/translation="NRVIVSPEAEEDILRSTEMLRNIDSGALDTRKICAKFSTLRD
LHCARXSYOMCVDTGACORCAERLIDVTEVNSLAHPLIEKVAIHFCKDEPELV
APITRKVYVLTGVAAAREAVKIMRILSKVLNLSLEMAFPYSPAMNAEMLDLRYT
PAKIRYDLDDVDTALNCEDEIHVYVTKEDENHLELOEDC"

Query Match 62.7%; Score 374.6; DB 14; Length 16527;
Best Local Similarity 76.7%; Pred. No. 1.2e-109;
Matches 458; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

ORIGIN
QY 1 ATGAGCTTATGTCGACGACGACCTTACGACCTGTGATACCGACGCTCTAGTCTA 60
DB 14604 ATGAGCTTATGTCGACGACCTTACGACCTGTGATACCGACGCTCTAGTCTA 14663
QY 61 AATGCTGTGACACAGAGCTTTATCTGTGTAAGTGAATAAATGTTGGTGCAGAAAGG 120
DB 14664 AATGCTGTGATAGAAAGCTGCTGTGGGAGAGCTATATAATGCTTGTGCAGAAAGG 14723
QY 121 GCTTCCTAACGAGGATAGAAAGTGTGGTGTCTACTCTTTAGGCACTCCGCGGAGA 180
DB 14724 GCGCCGACCGAAGATTGAAACAGCTTTGGGCTACTGCTTACGCTCTCGAGAGAGA 14783
QY 181 ACCAGCTCTCTAAGGTCAGCGCGAGATTACAGCTTATATTTTCAATAGTTTCGGA 240
DB 14784 ACCAGCTCTCTAAGGTCAGCGCGAGATTACAGCTTATATCTTCAAAACCTTACGCT 14843
QY 241 GAGAGAAATGTGTAGTACAGAGGTTACCTTAAGAGTACTCGACGGGTGTGGCCT 300
DB 14844 GAGAGAAATGTGTGTGCGAGAGAGTCTCAAGAAAGTGTGAAGGGTGGCAGCCT 14903
QY 301 CTCACTAGGTTCACTAATACTAGAAAGTGTGGTGTCTTACTAGAGCTTACGTT 360
DB 14904 CTCACTAGGTTCACTAATACTAGAAAGTGTGGTGTCTTACTAGAGCTTACGTT 14963
QY 361 GACTTTGTATCGGATTAAGCAAAATTAACCCCACTCAAGCGCGCGGAGATTGGGG 420
DB 14964 GACTTTGTATCGGATTAAGCAAAATTAACCCCACTCAAGCGCGCGGAGATTGGGG 15023
QY 421 ATTCCGCTCAAGTGTCTTACTAGTGCAGATTTTCTGGGTACTTCCCGGAGCTCTCT 480
DB 15024 ATTCCGCTCAAGTGTCTTACTAGTGCAGATTTTCTGGGTACTTCCCGGAGCTCTCT 15083
QY 481 GAATTACAGAAAGTACGAGATGTTGCGAGATGTAAGCTCTAATAAAGTGAAGGTGA 540
DB 15084 GAATTACAGAAAGTACGAGATGTTGCGAGATGTAAGCTCTAATAAAGTGAAGGTGA 15143
QY 541 GTGCTAATATACCAAGTACGAGATGTTGCGAGATGTAAGCTCTAATAAAGTGAAGGTGA 597
DB 15144 GTGCTAATATACCAAGTACGAGATGTTGCGAGATGTAAGCTCTAATAAAGTGAAGGTGA 15200

RESULT 6
LOCUS BYU51931 10545 bp RNA linear VRL 05-Apr-1999
DEFINITION Beet yellow stunt virus helicase gene, partial cds; and RNA
genes, complete cds.
ACCESSION U51931 L20761
VERSION U51931.1 GI:1388128
KEYWORDS
SOURCE Beet yellow stunt virus
ORGANISM Beet yellow stunt virus
VIRUSES; ssRNA positive-strand viruses, no DNA stage;
Closteroviridae; Closterovirus.
REFERENCE
1 (bases 2734 to 5576)
Karasev,A.V., Nikolaeva,O.V., Koonin,E.V., Gumpf,D.J. and
Garnsey,S.M.

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
FEATURES
SOURCE
COMMENT
On Jun 25, 1996 this sequence version replaced gi:507953.
Location/Qualifiers
1..10545
/organism="Beet yellow stunt virus"
/mol_type="genomic RNA"
/specific_host="Sonchus oleraceus L."
/db_xref="taxon:35290"
<1..2133
/function="replication"
/note="ORF1a; similar to helicases of other
closteroviruses"
/codon_start=1
/product="helicase"
/protein_id="AAC55658.1"
/db_xref="GI:1388129"
/translation="SSPLCYDMLRSGYCPISNKYFSGVSDSVYFKAVPTINRYKRP
TLDSIFEVHVNHNVEYQKFLFEEGFGFINSGSGASVGEENFEERDKNKD
RESGTSSTRANFVNDVABGKTEKRSRGSDHSLKAPRICYGDREPRDPRK
TLVSKFLNLTENAFPEPMITEADLASVREFFYLDELPEIFLNKLRPSSE
VVEEDRFLICRAGDDARLLVRESDASLYGQDRPLDMEYEVEFTGULPNDEK
SGNKLPHANTKFIANSFELNSNSYRNFIENDSCRLTEAPGCGKFTLLASFYK
MKNKRLIVLVANSOVELIKINDSKHEKTKLLEKPAAGKRENTPSADSNV
TIDSVLMHIGTCKDVTVECEFWHAGAVTAVNTLVPSVFPVPSRGTHIERNE
YDVASFSDRLVAAKRIYQGVSYRPMVVCGLSKGHYPVTAATTVSEGSLLTI
TEINSVDVVASKATYTLFLOSKEKELEKRLAKKYKATYKTHENQDPTIDVTVY
RTKQEDAPFSSFNHINVAITRTESLTVAALARRVDNIAALCANELVDKRLTP
HSFGQSVLNIIVEPVYTDNSCKASAPINSINDLEDDVPGSTLSNFGDTSAMEQ
PFESGANNVTVRDSKPGSGYDHDEQV"

CDS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
FEATURES
SOURCE
COMMENT
On Jun 25, 1996 this sequence version replaced gi:507953.
Location/Qualifiers
1..10545
/organism="Beet yellow stunt virus"
/mol_type="genomic RNA"
/specific_host="Sonchus oleraceus L."
/db_xref="taxon:35290"
<1..2133
/function="replication"
/note="ORF1a; similar to RbRp's of other closteroviruses;
the BYSV RbRp is presumably expressed via a +1 ribosomal
frameshift"
/codon_start=1
/product="RNA replicase"
/protein_id="AAC55659.2"
/db_xref="GI:4559421"
/translation="SYVPSOAIIPKRPBLOENTLSYERNVNFITCFRSGPHYGRRA
MAAVYKCFPLDKYAIRSDITAITGSGFEMLSKSPGSKSLASITSEPLALEE
IVRFELTVNRDARKVLDASCLVQKIPPAOTLNFHKAALNATISQCFDEKRYVLSCLP
NIIFELTVNRDARKVLDASCLVQKIPPAOTLNFHKAALNATISQCFDEKRYVLSCLP
GDEPDLVVMWGEYRSEKNTMDQVGTVCQKSGASNTWINSVVTGLIILMYD
VSKFQALFVSGDDSLIFSADEIANYAEDICLEGTEFTLPSVYFCSKLVFGDK
CVFVDPYKLVKASGRSLSEDFELFVFSFDLTFEFGDERLVNTLSLVHLYKE
FEMGNTKALSTIHLRENTLSFSLFVKRIGWVYVVGAKAYILKKFLGNIIEPITTT
EGDANFYVYK"

```

/codon_state=1
/product="p30"
/protein_id="AAC5560.1"
/db_xref="GI:138811"
/translation="MALYKYSALALVDPEDRTTSPGFNLAFSACDARFOQGVSG
GYSVALTLPILFISVVEVOSNEPFLSKRYMNDERRIRITGEMTIPISDKLIFHVK
VKQIQGNGNLIVYNGPHILGGITIVRFDLQGCENAAADIACHTSFSEVSDVAD
CDPMSEIILINTSDNNRACDADAYRRVAFVEGDLDTVDVDSNSERVEVDLRAEE
SSSGDDPDVDVIVSGEPKSKYLRRNRIMCFLVLLIGATIVYPMISRGVGS"
4401..4456

```

```

/function="unknown"
/note="ORF3: similar to small hydrophobic proteins of
other clostereoviruses"
/codon_start=1
/product="p6"
/protein_id="AAC55661.1"
/db_xref="GI:1388132"
/translation="MDCILRAFLPGEFALVICFILAAYFAFPVQVTHSDDDVDVLI
RQSDLLAG"
4574..46409

```

```

/funcion="unknown.
/notes="ORF4; similar to HSP70 homologs of other
closteroviruses"
/codon_start=1
/product="p66"
/protein_id="AAC55662.1"
/db_xref="GI:1388133"
/translation="MVFQGVDFGTFPSSVCVFNSSGRHLVHKQKNSAIYIPCTCLFPLSPD
MASFGDDEETASLDIPNVKGGPFLRKWCGCEBNIEKSKLKHYSVILNSNGKGG
SRIPIFGYSGSGVSGVSGSLGGLLAFLPLQVLSAAL1EPFCETELIVSVDGCMQK
RLETEKCAKLSPEPTCGCHMMNBPASAALSTGDRMBANLISLVRCGGTSPVLSYD
NORFTVPCAGCMGPTGCGDVPBAKATVYOMANIPBEPANLISLVKESITLIDVPIY
TYKTKGSEKTYVNSRGCLAEVYIPVDPTIKYKGEVPELYIKNNALDPAKAKSIVT
VGGSSILPELKILSVDPFSECLIDLPDPRAAIVAGCALSSCLSESMKLVFQKRV
NLSIPIPTVBSIVLVPAQAPVFEVGRININASCVSGSISVLPBGRKTKLVNKKRY
FSQIVLGLKGLVGTDPPTIRIVTLATLTERSVSGVTKATIGLSAKRVFVGPADFSKR
EVSIVSITLHLDNQRNALLATLTLTKRREKQFSKSEQHLDLSNLDLPEESKRR
PNQYNEQIDVACVRIILKSGVQKTLRGARVEELSYRNIYEVQTIKI"
6285..7889

```

```

/function="unknown"
/note="OnF5; similar to analogously positioned proteins of
other clostroriviruses"
/codon_start=1
/product="p51"
/protein_id="AAC55663.1"
/db_xref="GI:1398134"
/translation="MSRRPTFAGSYWGSILFRHYGEPEWWSYLTETSMKRYPLKSESIT
TPDASLISASLRPRASGAEYELALLIFSDSITKSELSRSYRGINDINHSYV
AEELELVTVYTGICGFTISAVESFMGSRASAAVEHCMSLNSCGE.LINPDRARFT
QLFPRKATLEAOAVNTSGSVSYLVYCLQILYNKSCNAGTCLMESYFS.FRDEPFT
QSHDLYDRRLPLNPLLTGLVYLDCLERNHYRGSYLNLDNFR.FPKOTYLPIMIDIPY
SMELVYLRPLPIDPEYELIKEVPTNSVDIANVSNKLYVSDGLNENSLALEKKK
LISILRNDGDEGALINAAFCVGYGTYTRAPKRYVYRPTVELDGSIKSPYWSGCE
LIPDEI.QXVPEDEGLRRFNAAKRGAEALTYPRKQVLSFPITRLNAPSKYVINDYF
KQNSLGLTPEKILITACQDMMCCQRISSYKAPLIDQKRGELNSAKYIRLPTN
TLVRLAEKCLLNDAPEWMTITLNR"
7932.8515

```

```

//function="unknown"
//note="ORF6: similar to analogous BSV coat proteins of
other closteroviruses"
/codon_start=1
/product="np25"
/protein_id="AAC56664.1"
/db_xref="GI:1368135"
/translation="MPPOGAELVYHNANSSLEVPSSTREKVEKFNNDHKTFKQ
VNNLINDELRVLVSGKTEKLTNLXALBEDIVHAFVFLIASVYSISPEVYKSS
YSSISDIORXKTVNDAMIPVOYVILASIKNNKNGIRAFCSALSGMYLSVARLGPFA
RSVSGKAGSGSEYLGADFLISTCPIMSDHDPALVALASRNALDSRAASQIDKRWYSI
YDKRGKVVY"
8551
/misc_feature
//note="transcription start site for coat protein
subgenomic RNA"
8617..9264
CDS

```

```
misc_feature      8551      /note="transcription start site for coat protein  
subgenomic RNA"  
CDS               8617..9264
```

CDS

```

/!note="CP; ORF7; similar to coat proteins of other
closteroviruses"
/codon_start=1
/evidence=experimental
/product="coat protein"
/protein_id="AAC5665.1"
/db_xref="GI:1388136"
/training="MAGNEDSGSDSASQWTAXDMFAPENFASATCINGENK
KXLFEEFSRVATDVESGIPITLGWTLAYALATSTSKIDIEKTPVLSAKIDAN
VATVEDIADFNSTLLKNYKKNKLRVFAETEEYLRFVROYKHILPNTARAKHGI
PDDVSYLADPNQTSNLLKHEQAVYLEGSMNAITSSQTTHESAVNLKLYGSSK"
9261 9743
/feature_key="protein"

```

```

//function=UNKNOWN
//note="ORF8"
/codon_start=1
/product="p18"
/protein_id="AAC55666.1"
/db_xref="gi:1388137"
/translation="MWYIESLQFVTPYKSKALVLTWQDAFIEHLLGVHSSPPYAL
WYHSGAALADDSVSPYKSLQDYLNSLIGYRTRMLRKPTSSLEHLYRATNAY
SIFRCNAPFLILFYTSKGPESMDAIVLYNGITGPVDYTHLSSIVESYGNITFAE
"
9746..10306

```

```

/function="unknown"
/notes="ORF9; similar to the 20-kDa protein of CTV and to the 21-kDa protein of BYV, two other closteroviruses"
/codon start=1
/product="p22"
/protein_id="AAC55667.1"
/db_xref="GI:1388138"
/translation="MKLLSDSYDYDINIAKKCINLEINLNSVPLESCIASYNNELIK
VNAQOSAPAIKNEKENLEPPDLNVTTTLEAVDILQILRAINEKPFSCSNQOV
KMLINLYNTHPSNRRPVPFQIOSFMKRGVITKSLQNLGYKKEVATINFKMA
LDLVNAGVYTFEENLIGENLESPFL"
10304..10545

```

/note="similar to 3' UTR's of CTV and BYV"

Query March	14.5%;	Score 86.6;	DB 14;	Length 10545;
Best Local Similarity	50.4%;	Pred. No. 3.2e-16;		
Matches 212; Conservative	0;	Mismatches 209;	Indels 0;	Gaps 0;

QY	48	CGCCCTAGTCTAAATGNTGCGACAAAGAGCTTTTATCTGCTGAAGTTGAAAAATGTT	107
Db	8721	CGCGACTTGTTGAATGTTGAAAAACAAGAAAGATTGTCGAGCAATTTTCTGTGAAGT	8780
QY	108	GCTGACAAAGGGCTCTTACGAGGGTATGAAATGAGTTCGGTCACTCCCTTAGCG	167
Db	8781	GAAAGCTCAGGATCTACTGATCCGGCAATCCCACTTTGGAAATGACGCTGACG	8840
QY	168	ACTCGCGGCAAGAACCAAGTCTCCTTAAGGTTACGCGCGAGATTCCAGACGTTATTTTC	227
Db	8841	CTTACTACTCTTAAGCACTTCCTTAAATTCGATTTGAGTAAGACACCGTTAGTAG	8900
QY	228	AAATGTTTGGAGAGAGAAATGATTAATACAGGGTACCTTAAGAGTACTCGA	287
Db	8901	CGCTAAGTATATCGGTCAACGTGACATTACTTACGAGACATTAAGAACTTCTGTAA	8960
QY	288	CGGAGTGCGGCTCTCACTAGATTCACTAATAAATTGAACGTTGGTCTGACTTTTAC	347
Db	8961	CTTCCTGACGTTGGTTTAAGAACTATTAAGATTAACCTCCGAGCTTCGCTCGCACTTTGA	9020
QY	348	TGAGGCTTAGCTTACATTTTGTATCCCGATTAGACAAATATACCCCAACTCAAGCGCG	407
Db	9021	GGAGGAGTACTTACGCTTTGTGAGGCAATACAAACATCTCCCGAATATGCAAGAGC	9080
QY	408	GGCGGAATTGGGATTCACGCTGAAGATTGTAATTAAGTGTGACAGATTTTCTGGTACTTG	467
Db	9081	GAAACAAACAGGATATCCCGCTGATTAAGTGAATTTAGACAGTCACTTCGCAACTAG	9140
QY	468	C 468	
Db	9141	C 9141	

/protein_id="AAFI4304.1"
 /db_xref="GI:6492372"
 /translation="MLAPRGRLIHFENTRDAMETFPNSDYLAESVFNKLNK
 ENDELIGVREPKSELVITDEDFVHLFALIRANITTSKVVYGAEXTIGK
 FLVKDANVPLIKECKKFNKPPVTPCATPEDYIVARSLPLFNRTIGKGI
 SGYELGADPLINTSVCLNDHEKATVLOASRAIDRAVSSVDGKIVSLFDGRIS"
 13641..14255
 /note="ORF6; CP"
 /product="major capsid protein"
 /product_start=1
 /protein_id="AAFI4305.1"
 /db_xref="GI:6492373"
 /translation="WSAEPIASAIFFENVSLADQCTLHGDCDKLRDPFECALIK
 VPEDKGLALGLYSCATIGTSNKVSVOPTSTFIASFGSELELTHEGLADFLDS
 OKLEGPNLRCFCRTFQDYISFAKEYGRFLPPIARANRGLPAEDHYLAADFLST
 STELTDLQGRLLARENATHTEFSSESPTSLIKOLRGALATGK"
 14252..14785
 /note="20 kDa protein; ORF7"
 /product_start=1
 /protein_id="AAFI4306.1"
 /db_xref="GI:6492374"
 /translation="MTSSVELAQTKPLFRVLLKGFVYIVAIETSESPVELPLVY
 LHLELVNKKGIESYIDFKSCMTLKSSVSYSRVSSGKSEDPSSSYSGKTPDS
 KIINRYTPRESGICLVFGMYGRDQRCVSEVLEWENFVGHGCTLPYCLNCELDK
 SGELELITFSKNEVFL"
 14769..15302
 /function="enhancer of RNA accumulation"
 /note="21 kDa protein; ORF8"
 /product_start=1
 /protein_id="AAFI4307.1"
 /db_xref="GI:6492375"
 /translation="MKFPNDCESTRALSSESLIRYVELGTNSPOSESEGINEN
 ELARFNHLVVEHERMEKHPKSSSLRTPSRIGELKRIAPLKVRYTPMKKTA
 SETLAPLEVCRIITGLTRDALREKRVKVSIVLPHESELKFEVTENPFTELLK
 LNTLSRVISQILGIAV"

ORIGIN

Query Match 11.8%; Score 70.4; DB 14; Length 15468;
 Best Local Similarity 47.5%; Pred. No. 6.1e-11;
 Matches 243; Conservative 0; Mismatches 266; Indels 3; Gaps 1;

26 TTAGCAACCTGCTATTAACGACGCTCTAGTCTAAATGTGTGACAGAAAGCTTTAT 85
 13678 TTGAAAGCTAGCTCGCAACCAAGCTTTTGACGCGTGAAGACTCGACAACTAC 13737
 86 CTGCTGAAGTGAAGAAATGTTGTGCAAGAAAGGGCTCTTAACGAGGCTATGAAGTGG 145
 13738 GAGAGGATTTGGAAGAGTGTGAATGAAGGGGTTCCGGAAGATMAACTCGTCTCG 13797
 146 TGTTCGGCTACTACTCTTACGACCTCGCGGCAAGACACGCTCTCAAGGTTACAGCGC 205
 13798 CGTTAGGACTTTGTTTGTACTCTGTGCAAGATAGTACTTTAATAAGTTAGTGTCC 13857
 206 CAGATTCAAGCTTATATTTTCAAAATAGTTTC--GAGAGAGAATGTGTTATACAG 262
 13858 AACCGACTTACTTCAATCAAAAGCTTGTTCGTGATGAGGAAGAAATGTTCTCACTC 13917
 263 AGGGGACCTTAAAGAGTACGACGGGTGCGCCCTCTCACTAGGTTCACTAATAAC 322
 13918 ACGGTGAAGTACGCTCTTTCGACCTTCAGAAACTTTAGAGGAAAGAGCTTAACAAAT 13977
 323 TTAGAAGCTTGGGTGCTACTTCACTGAGGCTTACGTTGCTTGTATCCGATTAAC 382
 13978 TCGGCTGTTCTGCGGCACTTTTCAGAAAGATTAATATCTTTCGCGAAGAAATACCGAG 14037
 383 ACAATTACCCCACTCAACGCGCGGGAATGGGATTCAGCTCAAGCTAGATTGTAAT 442
 14038 GAAAGCTGCTCGATTGCTAAGCCACCGCAAGGCTTACCTGCTGAATTAACATCACT 14097
 443 TAGCTGCAATTTTCTGGTACTTGCCTCCGCAAGCTCTGGAATTACAGAAAGTGAAGA 502

DB 14098 TAGCTGCTGATTTTCATATTCACATCAACAGACTTACTAGCTACAAAGAGTGTCTGC 14157
 QY 503 TGTTCGAGATATGATGAGCTCTTAAAACTGAA 534
 DB 14158 TGTTCGCGCGGGAAGAAAGCACTACACAGAA 14189

RESULT 8
 AF056575 15468 bp RNA linear VRL 02-JUL-1998
 LOCUS Beet yellows virus Californian isolate, complete genome.
 DEFINITION AF056575
 ACCESSION AF056575
 KEYWORDS AF056575.1 GI:3283076
 SOURCE
 ORGANISM
 Beet yellows virus
 Beet yellows virus
 Viruses; ssRNA positive-strand viruses, no DNA stage;
 Closteroviridae; Closterovirus.
 1 (bases 1 to 15468)
 Peremylov, V., Hagihara, Y. and Dolja, V.V.
 Genes required for replication of the 15.5-kilobase RNA genome of a
 plant closterovirus
 J. Virol. 72 (7), 5870-5876 (1998)
 2 (bases 1 to 15468)
 Dolja, V., Peremylov, V. and Hagihara, Y.
 Direct Submission
 Submitted (31-MAR-1998) Botany and Plant Pathology, Oregon State
 University, 2082 Cordley Hall, Corvallis, OR 97331, USA
 Location/Qualifiers
 1..15468
 /organism="Beet yellows virus"
 /mol_type="Genomic RNA"
 /isolate="Californian"
 /db_xref="taxon:12161"
 108..9381
 /gene="ORF1a/b"
 join(108..7983,8045..9381)
 /gene="ORF1a/b"
 /codon_start=1
 /product="fusion protein"
 /protein_id="AAC5115.1"
 /db_xref="GI:3283077"
 /translation="MAFLNSAVPSGCAAPAPFAHAGASPIVPSDFPCVPSPDISH
 RLTLSDFSVPRPLFNARVHLRASTNDPSPSLPLGHATFVLELNGSSAPPSIB
 RHIDVNRPSVPEFTVLVSLSLTSPRLALCDFLYCSKGPVEIASFSTPP
 CLVSNCAQIPTAHAMESIPEPTKTLPLAGRLQFHKRYTRPELTIHESGLAKTS
 ALGVSKNSRPTLVKSASGERKEAYEISRLQFERSRRQTPRSHRPKINKAVE
 PPPPEPRKKRKRASLPTKDEGTLTGTRPLSEPRKEEPLPKREVEIPVVK
 HAPAVSKPRTPRPVATGAEVYNAFPTCSRPRNPLIRASAYTFGFKMPLOF
 MEKQDYVYKSKSVSSCSVTSKPLALTSILKILPRISYNSERLKPYDHFIDGFEI
 EYHPLRGKLSVLLLPKGEAYCVTAATPQHAALTIRKQPSALKRPLVGRQVRS
 CYLAHALCALQKRTFREDFPVGMPTKVFAPALTEKQPSALKRPLVGRQVRS
 LEHCVASAFSSPFLPRFVGESEBPELTSIKHAISSYVERVSIHONDLARS
 VEKQIDPDEKESLSEKERSYVFPVGEAVOSGLTRAYQPFNLSFHSYSDHPAA
 AGSLLENETLSMAKSSSDIGGCLPHIKRGSIDYVCRPIYDKMAQORVSEIQ
 AGUYENLSREGLVAAQARVAVPCPTTLCNCKVSLVIMOVYDYMIAASAVLK
 SVAVLVNVTPELDERBARIDLCDDVYVDRPDVQVKGSSCCYCHLNTKNI
 MLTPATFTSGNLFSEVTEENRGVYIKITNSAPSEIRGYTLAPRACREYVYU
 PRFDKTLKFLSGDYIYLDAKFSVRVAVNSVNSVNSKTFEWMYSITASKSRVY
 ISGKVIHVDHIDLKHSBCFAVMAVAVRSRTTTEFLAKUNLYTGAACETIRFL
 PEMGRHAYAEINRSFRKMSIILSAGLYEFLDNLQHLLEYSEVAVRSIANG
 EYDCNENRVLTEITIAADKRSIAQSGALSVPOTOPRGGLRSGSRGVSFLN
 VREGNLTPSGVDVAVRFLVKRFTPSDSIPVVMFPLTKEAASPFAVSVALMLR
 BAVSAFSSVNAARTYSEVYKTFVNTVRFLNFMASKTLTKKFFRPLASALATV
 RAKVTLSEVMEVSEILSDGERSAYBFCSYITLLTNGGRLLDPSPSAVITV
 LLDLTKISILEVLLQISVDSTASALIRVLSLISNFRMGHGIPTYKFLICG
 LVFVARKCVALCPGDMATYARFLYGVDDLEFLGRVNSIYVLCVAAGVDSIV
 SVLKISGVAKERVYGFSGKIKNLFNFRKAVYTRSSDLSDEDFSCDSKIDG
 LGGSRFTLSLLDIFENFLSKLVLENACFSAYERTIERMKLYFFPLNLSBEAR
 RLIRGAGFDYVSDSAFDEMDLROAFVLEVSDESVTYDQKPTVLSLVANSRFL
 EFCNCPKRFVAVSYFKALYSRLRLRYLPMWRNLSDFGLGNEKALLAFLKTCV

FEATURES

source
 gene
 CDS

mat_peptide
 /gene="ORF1a/b"
 /product="papain-like leader peptide"
 /function="autocleavage"
 /function="replication enhancer"
 1872..6233
 /gene="ORF1a/b"
 /product="putative methyltransferase and RNA helicase"
 /note="ORF 1a"
 /codon_start=1
 /product="6.4 kDa protein"
 /protein_id="AAC25116.1"
 /db_xref="GI:3283078"
 /translation="MDCVLRSYLLAFGFWICLFLCLVFIWVYKQILFNTPEPSN
 EARFNRSTVY"
 9445..9609
 /function="unknown"
 /note="ORF 2; ps; small hydrophobic protein"
 /codon_start=1
 /product="6.4 kDa protein"
 /protein_id="AAC25116.1"
 /db_xref="GI:3283078"
 /translation="MDCVLRSYLLAFGFWICLFLCLVFIWVYKQILFNTPEPSN
 EARFNRSTVY"
 9609..11405
 /function="unknown"
 /note="homolog of HSP70 proteins; ORF 3"
 /codon_start=1
 /product="65-kDa protein"
 /protein_id="AAC25117.1"
 /db_xref="GI:3283079"
 /translation="MNVFGLDRTGTFSSVCAVYGERVYLFKORDAVYIPFVFLHSDT
 QVAFRTDAVYLNDSVRGRTYRDLKRMVGCPEENVGYLTKKHVTELLKVAOS
 SKSTVLDCTSGVLPQNALPGLIATFVRLISTASBARCQCTGTCVCPANYNLO
 RSFTESCYNLSGFCVYVWNEPSAALASASRIKATISPLVDFGSGVSVSAL
 NNTFVVASGDMNLGDRIDKAFVHLVKKQLPNYKIDISFLKESJKVSVLNF
 PVTSEONKVDVNVSELAEVAPFVEERTIKIVVEYKRSNMLSESVAKLMTV
 GASSYVGLISLVSVPFECLVLPDARAVALGALVSACLNDSFVCLVCAHN
 LSTSKYCESIVCPAGTPIPTGVRTVMNTGASVYSAALFEDGPVCLNKRKIF
 SGVVALGNVSSNTRTVPILTLEINVSVGTITPILVGTGPIKLVGNAAVDESSY
 QLGKRVADJHKNSDKVLIHAIITTPFKKLTLEGDRAFLIKRLSADYRBAKFS
 SYDDAVNSSELLGRVLPKILRGSREKLDV"
 11311..12972
 /function="unknown"
 /note="ORF 4"
 /codon_start=1
 /product="64 kDa protein"
 /protein_id="AAC25118.1"
 /db_xref="GI:3283080"

CDS
 /translation="MTTRFSTPANYMGELFRFFGGRKMLMSAASVRRPYDSD
 PRESNGEVSRRKNSDSTGESFYREERSLITLTPKTEYVCLGVAMEALSGNRLSD
 YNSELNVVDYKRVGCKPNIQYTERVRLNGLAEPSVHEVCHVMSNSGELINRDK
 TKRFVSLIRFGKQVSTDEAVSSYLDYLSHCLNLYXTCNLSNSGKALYDELK
 YVLYLETSDELSLNDPLVAGVYDMCFEETILKSYLKNKRLYESTFEDSN
 SEIFSMQNEPADVRLLELDTTELLKIPITNIDSTFLKNKRLYESTFEDSN
 ELIKVDSLLTDNDDELKLAQRMVGHAYGVFRAQKRVRADEYLPALGFT
 INNSGVEEPDELOKMPSVYRRFCGSLSEHAFSIIFRFGVFPPIRLNVVYYS
 YLVVDYRHYKRAGLTODELLISNIEPVAECVCEVVALQARAQGEKEFGQWG
 VKNVEPBRASSIRVKNNSILNTVMKQVARGRLNPLRHKH"
 12920..13570
 /note="24 kDa protein; ORF 5"
 /codon_start=1
 /product="minor coat protein"
 /protein_id="AAC25119.1"
 /db_xref="GI:3283081"
 /translation="MLAPRGDGLIHFTNTDPMETPENSVDLAEYSEVNPKNLRK
 ETBBLGVTERKSELVITDEPVGHLAALIRANITSTTVNVVAGVAYTIGGK
 FLVDAVAFPLIECMKRNKENVPTFCATFEDAYIVARSLPKFLARTIGKRIIP
 SGVEFLADFLITVCLNDHEKAIVLQASRAIDRAVSSVDKIVSLFDLGRLS"
 13641..14255
 /note="22 kDa protein; ORF 6"
 /codon_start=1
 /product="major coat protein"
 /protein_id="AAC25120.1"
 /db_xref="GI:3283082"
 /translation="MGSABEISAIATFENSVLDQCLHGECDKLRNFEELKLG
 VPEBKLGIALGLCYSCATIGTNNKSVQPTSPFIASFGSEGLTTHLELSFDS
 OKLEGGKRLRCGRTFOYISFAEYGRPIPARARHLPADHDYLAADFIST
 STETDIOGGRLLIARNATHTEPSSSPITSKQIGRLATOK"
 14252..14785
 /function="unknown"
 /note="ORF 7"
 /codon_start=1
 /product="20 kDa protein"
 /protein_id="AAC25121.1"
 /db_xref="GI:3283083"
 /translation="MTGSVELAOTKRPVLLIKGFVYVIALETESEPEVELPY
 LHDLELVNKKGLIESYIIPKSCMTLRKSSVSYTRVSGKSEDEPSSYSGKTPDS
 KIILNRKYTFEESGIQVFGMYGRDQVSSBYLMEVNVGARGTLPCLNCELDK
 SGSELRLITFSKNEVFL"
 14769..15302
 /function="replication enhancer"
 /note="ORF 8"
 /codon_start=1
 /product="21 kDa protein"
 /protein_id="AAC25122.1"
 /db_xref="GI:3283084"
 /translation="MKFFPNDCTSRASISSESLRRVKELGINSPOSEVSECINEFN
 ELARFNLVTVHEREMEGHKPQSSSLRTPSRIGEMLKEIRAFKRVVTPMKIETA
 SETLNAFLFEYCRITGLTRDALARKKRVKRSYVLFHSESLKEVENTENFSETELIK
 LNLSTRVISOGLIAIV"

CDS
 /function="unknown"
 /note="ORF 9; small hydrophobic protein"
 /codon_start=1
 /product="6.4 kDa protein"
 /protein_id="AAC25116.1"
 /db_xref="GI:3283078"
 /translation="MDCVLRSYLLAFGFWICLFLCLVFIWVYKQILFNTPEPSN
 EARFNRSTVY"
 9445..9609
 /function="unknown"
 /note="ORF 2; ps; small hydrophobic protein"
 /codon_start=1
 /product="6.4 kDa protein"
 /protein_id="AAC25116.1"
 /db_xref="GI:3283078"
 /translation="MDCVLRSYLLAFGFWICLFLCLVFIWVYKQILFNTPEPSN
 EARFNRSTVY"
 9609..11405
 /function="unknown"
 /note="homolog of HSP70 proteins; ORF 3"
 /codon_start=1
 /product="65-kDa protein"
 /protein_id="AAC25117.1"
 /db_xref="GI:3283079"
 /translation="MNVFGLDRTGTFSSVCAVYGERVYLFKORDAVYIPFVFLHSDT
 QVAFRTDAVYLNDSVRGRTYRDLKRMVGCPEENVGYLTKKHVTELLKVAOS
 SKSTVLDCTSGVLPQNALPGLIATFVRLISTASBARCQCTGTCVCPANYNLO
 RSFTESCYNLSGFCVYVWNEPSAALASASRIKATISPLVDFGSGVSVSAL
 NNTFVVASGDMNLGDRIDKAFVHLVKKQLPNYKIDISFLKESJKVSVLNF
 PVTSEONKVDVNVSELAEVAPFVEERTIKIVVEYKRSNMLSESVAKLMTV
 GASSYVGLISLVSVPFECLVLPDARAVALGALVSACLNDSFVCLVCAHN
 LSTSKYCESIVCPAGTPIPTGVRTVMNTGASVYSAALFEDGPVCLNKRKIF
 SGVVALGNVSSNTRTVPILTLEINVSVGTITPILVGTGPIKLVGNAAVDESSY
 QLGKRVADJHKNSDKVLIHAIITTPFKKLTLEGDRAFLIKRLSADYRBAKFS
 SYDDAVNSSELLGRVLPKILRGSREKLDV"
 11311..12972
 /function="unknown"
 /note="ORF 4"
 /codon_start=1
 /product="64 kDa protein"
 /protein_id="AAC25118.1"
 /db_xref="GI:3283080"

ORIGIN
 Query Match 11.0%; Score 65.6; DB 14; Length 15468;
 Best Local Similarity 46.9%; Pred. No. 2.2e-09;
 Matches 240; Conservative 0; Mismatches 269; Indels 3; Gaps 1;
 QY 26 TTAGCAACCTGTGATTAACGACGCTCTAGTAAATGNTGTGACACAGAAGCTTTAT 85
 DB 13678 TTGAAGAGTGAAGTCTCTAGACCAAGTGTGTCACGATGAAGACTGCACAACTAC 13737
 QY 86 CTGCTGAAGTGAAGAAATGTTGTGACGAAGAGGCTCTTAACGAGGATTAAGAGTGG 145
 DB 13728 GGAGGAATTTGGAAGAGTGTGAAGATGAAGGGGTTCCGGAAGATTAACGCGTCCG 13797
 QY 146 TTTTGGGTCTACTCTTTACGACCTCGCGGGAAGAACACAGCTCTTCAAGTTCAGCGCG 205
 DB 13798 CGTTAGGACTTTGTTTACTCTCTGTCGACGATAGTACTTAATAAAGTTAGCGTCC 13857
 QY 206 CAGATTCAAGAGCTTATTTTCAATAGTTTGC--GAGAGAGGAATGTGTAGTATACAG 262
 DB 13858 AACGACGCTACTTCAATCAAGCTTGTTGCGTAGAGTGGAGGAATGTTTCTCTCACTC 13917

QY 263 AGCGTGAACCTTAAAGAGTACTGACGGGNGGCCCTCTCACTAGTTCACATAAATAC 322
 Db 13918 ACGGCGAAGTGAAGTCTCTTCTGACTCTCAGAACTTTAGAGGAAAGCCTTAACAAAT 13977
 QY 323 TTGAACGTTGGTGGTCTTCTGACTGAGGCTTACGTTTGTATGCGGTATMAC 382
 Db 13978 TGGCGGTGTTCTCCCGCACTTTTCAAGAGATTAATCTTCGCGAAGAAATACCGAG 14037
 QY 383 ACAATTAATCCCACTCAACGCCGCGGCGAATTGGAGATTCAGCTGAAGATCTGACT 442
 Db 14038 GAAGACTGCTCCGATGCTAGAGCCACCGTACGCTTACTCTGCTGAAGATCACTACT 14097
 QY 443 TAGCTGAGATTTTCTGGTACTTGGCCGAAAGCTCTCTGAATTACAGCAAGTAGAGAGA 502
 Db 14098 TAGTGTGATTTCAATGACATCAACAGAACTTACTGACTCAACAGGTCGTCTGC 14157
 QY 503 TGTTCGAGATGTATGAGCTCTTAAAAACTGAA 534
 Db 14158 TGTTCGCGCGGAAAGCCACTCAACAGAA 14189

RESULT 9
 LOCUS BYVCPG 2724 bp ss-RNA linear VRL 02-AUG-1993
 DEFINITION Beet yellows virus coat protein gene, complete cds, and two ORFs.
 ACCESSION M59452
 VERSION M59452.1 GI:323237
 KEYWORDS coat protein.
 SOURCE Beet yellows virus
 ORGANISM Beet yellows virus
 Viruses; ssRNA positive-strand viruses, no DNA stage;
 Closteroviridae; Closterovirus.

REFERENCE 1 (bases 1 to 2724)
 Brunsfeld, J., Moseley, J., and Hull, R.
 Nucleotide sequence of cDNA encoding the coat protein of beet yellows virus
 Unpublished (1990)
 JOURNAL Original source text: Beet yellows virus cDNA to genomic RNA.
 COMMENT Location/Qualifiers
 FEATURES
 SOURCE

CDS

1..2724
 /organism="Beet yellows virus"
 /mol_type="genomic RNA"
 /isolate="Broom's barn"
 /db_xref="taxon:12161"
 <1..1287
 /codon_start=1
 /label=ORF_1
 /protein_id="AA172953.1"
 /db_xref="GI:323238"
 /translation="PDEVKRLNGNMAEPLVHEHWSLSNSCGELINPKDTKRFVSLIF
 KGVVSTDAVAVSSYLIDLSHCLNLYEVCNLSNSGKALYDEFLKVIAYLEIS
 DLEVSPSDNLTAVGVIYDMCFEYNTLSKNIKSFDFLSLYLLEISPMWSE
 OPADYRLIFLIDPRAELIKIPLINTIDSTFLNKRKIESYEDNSNLIKVDVS
 LITDNPBLKQKRWGCHCYGFRIRQKRYKRDVAYLPLAGFTINMGVVEH
 FDELQKMPVSVRPFSGSLSHAFSPKRGVGFPPTRILNVPKYSILINDYTH
 VKRAGLTDELITLISIEFDVAMCCEREVALQARRAQKPEPQGMKGVANVSPRA
 RSSIRVKSNSLSLITLIMKDVAGARQGLNLHKKH"
 1235..1885
 /codon_start=1
 /label=ORF_2
 /protein_id="AA172954.1"
 /db_xref="GI:323239"
 /translation="MLAEGRGDLIHFTENTRDAMETFPNSYDLAESEVNPKNLAK
 ETDELGIIRERFSELEVIDEDFYKHLAFALIRAGNTTSKVNIVGAYEYTGKPK
 FLVDAMVPLIKKCKKFNKPNPVTFCAITFEDAYIVIASLPRKLNIRITGRGS"
 SGEFLGADFLVATSVCLNDHKAIVLQASBAIDRAVSSVDKIVSLFDLGRIS"
 1956..2570
 /codon_start=1
 /label=ORF_3
 /product="coat protein"
 /protein_id="AA172955.1"
 /db_xref="GI:323240"

CDS

ORIGIN

Query Match 10.8%; Score 64.6; DB 14; Length 2724;
 Beet local Similarity 46.8%; Pred. No. 3.5e-09;
 Matches 239; Conservative 0; Mismatches 269; Indels 3; Gaps 1;

26 TTAGCAACTGTTGATTAACGAGCGCTCTAGCTTAAATGCTGACAGAAAGCTTTAT 85
 Db 1993 TTGAAGAGTACGCTCGACAGCAACGCTGTTGACGGTGAAGACTGCGACAACTAC 2052
 QY 86 CTGCTGAAGTTGAAAAATGTTGTCAGAAAGGGCTCTTAACGAGGATAGAGTGG 145
 Db 2053 GGAAGAAATTCGAAAGAGTGTGAAATTTGAAGGGGTTCCGAGAACAACTCGTCTCG 2112
 QY 146 TGTTCGCTACTCTCTTACGACTCGCGGAGAACACAGCTCTTAAGGTTACGCGC 205
 Db 2113 CGTTAGAGCTTTGTTTGTATTCCTGCGACGATAGTACTTTAATAAGTTAGTGTCC 2172
 QY 206 CAGATTCAAGCCTATATTTTCAATATGTTTCGAGAGAGAAATG---TGTAGTACAG 262
 Db 2173 AACGACCTTACTACTTATCAAAAGCTTCGTGGTGTGGAGAAATGTTCTCTACTC 2232
 QY 263 AGGTGACCTTAAAGAGTCTCGACGGGNGGCCCTCTCACTAGGTTCAATATAAC 322
 Db 2233 ACGTGAACCTAGAGTCTTTTCTGACACTCGAAGCTTTGAGAGAAAGCTTAAAGT 2292
 QY 323 TTGAACGTTGGTGGTCTTCTGACTGAGGCTTACGTTTGTATGCGGTATTAAG 382
 Db 2293 TGGCTGTTCTTGGCCGCACTTTTCAAGAGGACTCATATCCTTCGCGAAGAAATACGAG 2352
 QY 383 ACAATTAATCCCACTCAACGCCGCGGCGAATTGGAGATTCAGCTGAAGATCTGACT 442
 Db 2353 GAAGACTGCTTCCAGTCTTCTGAGCGAACCGTACGCTTACCTGGAATCACTACT 2412
 QY 443 TAGCTGAGATTTTCTGGTACTTGGCCGAAAGCTCTCTGAATTACAGCAAGTAGAGAGA 502
 Db 2413 TAGCTGCTGATTTCAATGACATCAACAGAACTTACTGACCTTCAACAGAGTGTCTGC 2472
 QY 503 TGTTCGCGAGTATGAGCTCTTAAAAACTGA 533
 Db 2473 TGTTCGCGCGGAAAGCCACTCAACAGAA 2503

RESULT 10

BYU71295

LOCUS BYU71295 612 bp RNA linear VRL 14-OCT-1996
 DEFINITION Beet yellows virus coat protein gene, partial cds.
 ACCESSION U71295
 VERSION U71295.1 GI:1619939

KEYWORDS
 SOURCE
 ORGANISM

Beet yellows virus
 Viruses; ssRNA positive-strand viruses, no DNA stage;
 Closteroviridae; Closterovirus.

REFERENCE 1 (bases 1 to 612)
 Creemer, R., He, X.H., Yang, C.H. and Grantham, G.
 Characterization of the 3'-proximal encoded proteins of beet yellows closterovirus
 Unpublished

JOURNAL
 REFERENCE
 AUTHORS
 TITLE

2 (bases 1 to 612)
 Creemer, R., He, X.H., Yang, C.H. and Grantham, G.
 Direct Submission
 Submitted (19-SEP-1996) Plant Pathology, University of California,
 Riverside, CA 92521, USA

JOURNAL
 FEATURES
 SOURCE

1..612
 /organism="Beet yellows virus"
 /mol_type="genomic RNA"
 /isolate="California BYV-CA"

```

/ab_xref="taxon:12161"
1..>612

```

```

1. >bl2l
/codon_start=1
/product="coat protein"
/protein_id="AAB17001.1"
/db_xref="GI:1613940"
/translation="MSGASGTSIAATFENVSLVDOTCHGEDGCDLRLNPEECILKLGK
VPEKCALAGICISGATGICGNKYVOTSPFIKAASGSEKFLTHGELRSLDS
OKLKEGKNGRLRCRCRFQCDYISPAKENGRLPIAIPANRGCLPAEDHYLAAPFIS
TELTLDYQGGELIARLNAAHTFSSDSPTSLKQIGRLATGD"

```

ORIGIN

Query Match	10.7%;	Score 64;	DB 14;	Length 612;
Best Local Similarity	46.7%;	Pred. No. 4.4e-09;		
Matches 239;	Conservative 0;	Mismatches 270;	Indels 3;	Gaps 1

QY	26	TTTGCACTCGGTGATTAACCGACGGCTTAACTCTCAATTAATGGTGTGCAAAACAACTTTAT	85
Db	38	TTGAAAACGTGAGTCTCGTAGCCAAACGTGTTTGCAACGGTGAAGACTCGGCAAACTAC	97
QY	86	CTGCTGAAGTGAAAAAATGTTGTGTCAGAAAGGGGCTCTTAAGGGGTATACAGTAG	145
Db	98	GGAGAAATTCGAAGAGTGTGAAATTTGAAAGGGGTTCCGAAAGATTAACCTGGTCTCG	157
QY	146	TGTTGSGTACTCTCTTTACGCACCTCGCGGCAAGAACAGCTCTCTTAAGTTCAAGCGG	205
Db	158	CGTTAGGAATTTGTTTGTATCTCTGTGGAGGATAGGATCTTCAATTAAGTTAGGCTC	217
QY	206	CAGATTCAAGCTTATATTTTCAAAATAGTTTC--GAGAGAGAAATGTTGTAATACAG	262
Db	218	AACCGACGTCTACTTTTCATCAAAAGCTTGCTGGTAGTGGAGAAAGAAATTTGTTCTCAGTC	277
QY	263	AGGCTACCTTAAGAAAGTACGACGGGTGCGCCTCTCAGTAGTTTCACTAATTAAC	322
Db	278	ACGGCAACTGAGTGCCTTTCTGAGCTCTCAAGAACTTTTAGAGGAGAAACCTTAACAAAT	337
QY	323	TTAGAAAGTTGGGTGTACTTCACTAGAGCTTACGTGATCTTTGTATGCGGTATAGC	382
Db	338	TGGGAGTTTCTGGCCGACCTTTTCAGAAAGATTAATATCTTTCGCGAAGAAATACCGAG	397
QY	383	ACAAATTAACCCCAACCAACGCGCGAGCGGAATTTGGGAGTTCCAGCTGAAGATTCTGACT	442
Db	398	GAGACTGCTCCGATTCGTAGAGCCAAACGTCACAGGCTTACCTGCTGAGAGATCACACT	457
QY	443	TACCTGCAAGTTTCTGGTACTTGTCCCGAAGCTCTGAAATTACAGCCAAATGAGAAAG	502
Db	458	TACCTGCTGATTTCAATGACATCAACGAACTTACCGACCTTCAACAAGAGTCGTCTGC	517
QY	503	TGTTCCGAGTATGTAGCGCTTAATAAACGAA	534
Db	518	TGTTGGGCGGAAAAAGCGCACTCACACGAA	549

RESULT	1
LOCUS	CLBYV3PH
DEFINITION	CleYV3PH 6746 bp RNA linear VRL 13-JUN-2002
LOCUS	Beet yellows virus genome 3'-proximal half, for capsid protein HS70 related protein, RNA-dependent RNA polymerase and ORFs.
ACCESSION	X53462
VERSION	X53462.1 GI:58878
KEYWORDS	capsid protein; heat shock protein 70 homologues; RNA polymerase; RNA-dependent RNA polymerase; unidentified reading frame.
SOURCE	Beet yellows virus
ORGANISM	Beet yellows virus Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae; Closterovirus. 1 (bases 1 to 6746)
REFERENCE	Agronomsky,A.A., Boyko,V.P., Karasev,A.V., Lunina,N.A., Koonin,E.V and Dolja,V.V.
AUTHORS	Nucleotide sequence of the 3'-terminal half of beet yellows closterovirus RNA genome: unique arrangement of eight virus genes U. Gen. Virol. 72 (Pt 1), 15-23 (1991)
JOURNAL	
MEDLINE	91116305

PUBMED	1990061	2 (bases 1 to 6746)
REFERENCE		Agranovsky,A.A., Boyko,V.P., Karasev,A.V., Koonin,E.V. and
AUTHORS		Dolja,V.V.
TITLE		Putative 65 kDa protein of beet yellows closterovirus is a
JOURNAL		homologue of HSP70 heat shock proteins
MEDLINE	91171285	J. Mol. Biol. 217 (4), 603-610 (1991)
PUBMED	2005613	
REFERENCE		3 (bases 1 to 6746)
AUTHORS		Kensung,S.A. and Maier,U.G.
TITLE		Phylogenetic analysis of the stress-70 protein family
JOURNAL		J. Mol. Evol. 39 (1), 80-86 (1994)
MEDLINE	94343547	
PUBMED	7545947	
REFERENCE		4 (bases 1 to 2677)
AUTHORS		Agranovsky,A.A.
TITLE		Direct Submission
JOURNAL		Submitted (22-MAY-1990) Agranovsky A.A., A.N. Belozersky
REFERENCE		Laboratory, Moscow State University, Moscow 119899, USSR
AUTHORS		5 (bases 2578 to 6746)
TITLE		Agranovsky,A.A.
JOURNAL		Direct Submission
REFERENCE		Submitted (27-JUL-1990) Agranovsky A.A., A.N. Belozersky
AUTHORS		Laboratory, Moscow State University, Moscow 119899, USSR
TITLE		Location/Qualifiers
FEATURES		1..6746
SOURCE		

CDS

```

/organism="Beet yellows virus"
/mol_type="genomic RNA"
/db_xref="taxon:12161"
<!--.659
(40403, 659)

```

CDS

```

710. 874
/note="unamed protein product; ORF2"
/codon_start=1
/protein_id="CA37550.1"
/db_xref="NCI_58879"
/db_xref="SPTREMBL_008542"
/training="MDCVLSVILLAFGLICFLCFLVVFIMFVKKILRPTTAS

```

CDS

```

3740      .2670
/codon_start=1
/product="HSP70 homologue"
/protein_id="CAA37551.1"
/db_xref="GI:58880"
/db_xref="GCA:P37092"
/db_xref="SMIS-PROT:P37092"
/translation="MVVFGGDFDPTGRTSSVCAVYGBSELYIFKRDSDAYIPYVFLHSPD
OEVAGADIAVLNDLSVRGDFGDFRTSSVCAVYGBSELYIFKRDSDAYIPYVFLHSPD
SKSTIKDQVCGVYDPONALPGJLAIFVALLSTABEAKCCCTCIVSPANNICL
RSFTESCYNLGGCVYVMEVAPPAALASCSRKATSTSVLYVDGSGGTFPVSYISA
NNTFWVRAGSDMLVGGRLDIFVFLNKKALPNVYKIDISFLKEBLSKSVFLNME
PVSGHGGVGVVYVNDLLEVAAPFVERIKIVKEYKCYSSNRLPPNVAKLPLNME
GGSSTPELISLRSSIFPDNCEVLDAAAVAGCAVSAALTRNDSFMLVDCAAHNS
LSISKVSGEYVYCPKASPIPTFGVCTVDMAGNSAAYSAALFGDGDQRLINKRI
FEDVAVLGGVSTGCTKIVTELPELINSVSGTISFSLVDPFVYKLTIGNAAVYDSS
DGEVAVNADLHKNSDVKVLIHALVYQPFQRKLTIDGKALPLKLTIDYIREAKP
SYDAEVNADLHKNSDVKVLIHALVYQPFQRKLTIDGKALPLKLTIDYIREAKP
2576      .4237
/note="unmerged protein product; ORF4"
/codon_start=1
/protein_id="CAA37552.1"
/db_xref="GI:58881"

```

CDS

```
/note="unnamed protein product; ORF4
/codon_start=1
/protein_id="CAA37552.1"
/db_xref="GI:58881"
```

/db_xref="SPTREMBL:Q06110"
 /translation="MTFRSTPTANYWGELEFRFFGGQGEWKNLMSAASVSRPYSID
 PPSGVLISRTKTEGSEFSEFSEFSLTLPKTYEYCKLQVAMALANGNRISD
 VNSFSEVLDVTVGCKENIQSVFVKKINGNAVPLVAGVLSNGSGELINPD
 TRFVSLIFKQDLAESEDEAVSSSYDYLSCNLTENLSNNGKSLIDEFLK
 HYDYLSEDLSEYSPSDNPLVAGILYMCCEYNTLKSQYKNISEFDFCSLYPL
 SEFPMNMRPADVRLFEIDAEALLKPTINMHSDTEFLYKRLYSEFSDSDN
 RLKYVDSILTRNDPDLKAGWGHCHYGVRTQTPVXRDAEYKLPALGPRV
 INMSGVEEFEBLOKMSISYRAFCOSLSHEAFSTFKGVSFPPTILNTPVKS
 YLNDVYRHRVKEVGLIDBELTILNIEFDVAMCCERVALQARRAREPRQMGK
 TNEISPHARRSIRVXKNDLSLNLMEVDVARSQRPLPLRRH"
 4185..4835
 /note="unnamed protein product; ORF5"
 /codon_start=1
 /protein_id="CAA37553.1"
 /db_xref="GI:58882"
 /db_xref="GOA:Q08536"
 /db_xref="SPTREMBL:Q08538"
 /translation="MLAPARGDLHFTENTRDAMETFPNSYDIAEYSEVNPKNK
 EIDELGVIRERKSELVITDEPVKHLAFLIRANITTVKTVGAYEYTGKPK
 PLVDAWVPLIKCKCKENKPNPRTCAFEADAYIVIASLPLRLNRTIGRGIP
 SGVEFLGADPLTATSVCLNDHEKAIYVLAQSAADIDRAVSSVDKIVSLPLGRIS"
 4906..5520
 /codon_start=1
 /product="capsid protein"
 /protein_id="CAA37554.1"
 /db_xref="GI:58883"
 /db_xref="GOA:Q08531"
 /db_xref="SPTREMBL:Q08531"
 /translation="MGSAPISALATFENVSADOTCLHGEDCKLRNFEELKLG
 VEDNGLALGCLVSCATIGTSKVNQTPSTKAFGGKXELVTHGNSFLDS
 OKLIEKPKRLKCFORTPOKDYISLRKYRCKPLPIANVHGLPADHYIADPFI
 SEITLQOSRLILARENATHTEFSSESPVSLKLGGLGTGR"
 5317..6039
 /note="unnamed protein product; ORF7"
 /codon_start=1
 /protein_id="CAA37555.1"
 /db_xref="GI:58884"
 /db_xref="SPTREMBL:Q08544"
 /translation="MTSSVELAQCKPLRVLILKGFVYVVAFTPEBSSEHPLVY
 LHDFTNIRKGRKIASYVDPMSCMTRLKPSVSTYRSKSSSDSISPSGTRRS
 KVLNRKVTFTENGVDVFGMYGLQRCVSDSYLFEVAFGAGHGLITLNCLELCK
 SGSEELTFPSKREVLKRM"
 6034..6567
 /note="unnamed protein product; ORF8"
 /codon_start=1
 /protein_id="CAA37556.1"
 /db_xref="GI:58885"
 /translation="MKFPLKDGTSRALSRSLSLRVKEICTNSQOSEISECDEN
 ELASFNLVTVHEHMEQHPNQSSKLVRPRIGENKKEIRATLKRVTPEHKEA
 SDLNAFLSEYCRITGLAREDLREKRRKVSVLFFHSELLKEVTENMESYLELK
 LMSLRVSISSQILGMAI"
 ORIGIN

Query Match 10.7%; Score 63.8; DB 14; Length 6746;
 Best local Similarity 48.1%; Pred. No. 7.4e-09;
 Matches 213; Conservative 0; Mismatches 227; Indels 3; Gaps 1;

ORIGIN

QY 95 TTTAAAAAATGTTGGTGCAGAAAGGGCTCTCTAAGAGGCTATAGAGTGTGTTGGTC 154
 Db 5012 TCGAAGAGTGTGTAATTAAGAGGGTTCGGAAGATTAACCTCGAATCGCTTAGAGAC 5071
 QY 155 TACTCTTTACGACTGCGGAGAACACAGCTCTCTAAGGTTCA---GGCGGAGATT 211
 Db 5072 TTTGTTGATTCCTGTGCTAGCAATGAGCACTTCAACAAGTTAACTCCACCGACGCT 5131
 QY 212 CAGACGTTATTTTAAATAGTTTGGAGAGAGAGAGTGTAGTAAACAAGGTTGACC 271
 Db 5132 CTACCTTCACAAAGTGTGTTGATGGAGGAGAACTGACCTCACTACGAGTGAT 5191
 QY 272 TTAAGAAGTACTCGACGGGTGGCGCTCTCACTAGTTCCTAATAAATTGAACGT 331

Db 5192 TGAATCTCTTCTGGGGCTCTAAAAAATTGAGAGGAAAACTTACAAATTCCGGTGT 5251
 QY 332 TCGGTGACTTCTCACTGAGGCTTACGTTGACCTTTGTATCGGTATPAGACAAATTAC 391
 Db 5252 TGTGGCTACTTCTCAGAGGACTACATATCTCTTCCCAAGGAATACCGAGGAAATTAC 5311
 QY 392 CCCAAGCTCAAGCGCGCGCGGGAATTTGGGAAATTCAGCTGAAGATTGTGACTGACGAG 451
 Db 5312 CTCCAGTTTCCAGAGCTTACCGCTCACGGTCTACCCGCTGAAGATCACTATTAGCCGCTG 5371
 QY 452 ATTTCTGGGTACTTGTCCCGAGGCTCTCTGAAATTACGAAAGTAGAGAGATTTCCGCA 511
 Db 5372 ACTTCATATCGACGCGCAGGAACTCACTGACCAACAAAGTCGTCTGTGACCGC 5431
 QY 512 GATATGACGCTTAAACCTGAA 534
 Db 5432 GCGAAACGCCACCTCACACGGA 5454

RESULT 12
 BYUAA
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 X73476.1 GI:405624
 Beet yellows virus (isolate Ukrainian BYV-U) complete genome.
 coat protein; heat shock 70-related protein; heat shock 90-related
 polymerase; RNA helicase.
 Beet yellows virus
 Viruses; ssRNA positive-strand viruses, no DNA stage;
 Closteroviridae; Closterovirus.

SOURCE
 ORGANISM
 1
 Agranovsky, A.A., Koonin, E.V., Boyko, V.P., Males, E., Frotschl, R.,
 Lunina, N.A. and Atabekov, U.G.
 Beet yellows closterovirus: complete genome structure and
 identification of a leader papain-like thiol protease
 Virology 198 (1), 311-324 (1994)
 94082464
 MEDLINE
 PUBMED 8259666
 JOURNAL
 TITLE
 AUTHORS
 REFERENCE
 COMMENT
 FEATURES
 source

2 (bases 1 to 15480)
 Agranovsky, A.A.
 Direct Submission
 Submitted (18-JUN-1993) A.A. Agranovsky, A.N. Belozersky Institute,
 Moscow State University, 119899 Moscow, Russia
 Related sequence: X53462.
 Location/Qualifiers
 1..15480
 /organism="Beet yellows virus"
 /mol_type="genomic RNA"
 /isolate="Ukrainian (BYV-U)"
 /db_xref="taxon:12161"
 /clone="BYV 936, x19, R9, R3, 36A, R2, 36, p43, 111-7,
 112, 115, 1213, 1210, 124, 1313, 1311, 142, 143, 1311,
 1311, 142, 143, 156, 1510, 154, 1518, 1520, 169, 1615, 175"
 1..107
 join(108..7983, 8045..9393)
 /gene="ORF1a/b"
 join(108..7983, 8045..9393)
 /gene="ORF1a/b"
 /codon_start=1
 /product="fusion protein of papin-like protease,
 methyltransferase, RNA helicase"
 /protein_id="CAA51871.1"
 /db_xref="GI:809539"
 /db_xref="GOA:Q08534"
 /db_xref="SPTREMBL:Q08534"
 /translation="MAFLNVSAYDSCAFAFAPAHGASIVDPSPFCVPRYDDISH
 RLTLISLDFSPRPLSLNARVHUSLSTNDPLSLPLGFAHETVLELNGSAFSLIS
 RHIDVNRPPSVFPLEVLSLRTSRLFALCOFLVCSKRGPCVLEIASFTSP
 ALGVSCVAQIITFAEMESIRFPTKLPAGFLPHKAKYTRPETLI IHESGLAKTS
 ALGVSCVNSRPIITVKSAGKEVAYISKDFERSRRROCTRVSHKRXKINKAVE
 PFFPEBPKKDKRRASLPTEDEGFTFGTLRPLSRTPEBPRLPKPREVEIPVVK

```

HAYAVATSKVTRFRPVATTGATYVNAARNCSSRPNNHPIIINASTIFPFKMPQORF
MEKKEVYVRSKVVSSCSVTKSPLEALAILNKLVOYSINSEKLFYOHIFGDEBEI
EVLHRLGGKSLVLLILPKGEAYCVVATATQYHAALITARGDPRVGLDQYRDEGL
CYLHAALICOLQRTFREDFEVGMPTFVAKRLTELGPSALHGHVRRQVRS
LFHCDVASAFSPYSLPRFVIGVEEAPETISLHKAIESYERVSIAHKNILARS
VSKLIDFKDEIKSLKSKRSVYVPMYMGAVOSGLTRAPQPNLSFTSASDHPAA
AGSRLIDENETLISMAKSPSDIGGCPFIHKRGSTDYHCRPIYDMKDQRVRSLEO
ARGVENUISREQLYEAQARYVCPHTLGNANYSVDLIVQYDASINIAAMTLKE
SKVAILNMTFEGELDEREAFIDALGCDVVTTRDMQVYFGSCCTHAKSNKSI
MLVAFPSGNLPSVEWYENRMGNVYKITSRAYSSEIGVTKLRPRACSTNVVVKL
PREDKTLKTFSGDYIYLDKAVSRFVDYVNCVNSKTFEWMVSIKSKSRVY
ISGVYHIDVHIDLHSECAVMAVAVASRTTEFLAKNLVYTGDSCEETIRFL
PREMSRAYAVIENSFRKLMSKILISADYEDFLDNLNOLHLEYSVEVRSIIONG
EVDGEMENRVLTEIABADRKSLAOGSLGSLSVPTOPRGLRGGRSRSGPVLN
VEEYGNTPFSVQDAVRFVIVKTPSPDSPIFRVYRFLDLAASAPFVSIVGLCMTR
EAVASRSMVADRVSSEVYKTFVNRVYKRLNMSAKTLTKTFRFLPSASLACTVY
RKAKVILEATWVWFESILSDSGEYSAVESCSVITLTLNSGRLLPGFSPALITVE
LIDLATKISIEVLKQISPADSTASALYRVLSEILSNFRMGHBGIFTKVPILLGCF
LPYVRKCVALCVPGDMATYARFLEYVDLJFLGRSVNSIXNYLCVVAAGLVDIYD
SVYLKSGVAKERYLGFESKIINKPLVFRKAVVTRTSSDLSDEYFSCDEKPG
LRGSSRPTSLRLDIPENFLKSKSVIENACSAVERLERMMKLYEFLNLSSEEAR
RLTRCAGDPYLSADFADEDEMLROAFEOVSSDSDSVYDGGKPYLBSYLVNARFL
ETPCAGSPFKVNSYKALYSRLVLTPEVDNISDSGLKGNKALAAPEFTCV
ITACEVSQICCLRLKLCWGTFRACGLVRLPTTYSSTRYLSRVVAVAVCPVLVNE
LDGLSDGLTNMGVSFRRLFVALRRALASVNSALRRKITEFLFGVHHPFVAVIET
NEVAPRLSEVDIDVDQDFGSDSESVSDSVANSPRGLGGRSRSNPLSLVAVY
FKLARIPLRLFLRLNFAYFERRLASKRLKTFIGLARFNFLSLTVVYLJOEDS
VLNAPLIDELILNLSGVNVLPLVSWRGSITLALAIYSGFASPLGMCORVDMC
SSSAGNCFMSPVRTKGFVPSSGSGTASVWERLEALSDIREHVLTCGVSGSE
RESKEVTEPGIEHTSEDVYPIRSHQPLSGGCSYSEDEBENRANLPHYSKIYSE
RLGLETARRMKRTHLHVSSEFINAINSNBOBRPIIYDHSBESALTNSEREYVYJEL
ALFPLSKRHYNDQLVANPNROECICDEDEMPVLIRAGQSVYSGRNRPLKTFKFG
HECFRSGGLVPYDQISRVDITPHIQTFNSAVALSGIUSITFTFTNLASVNLAYE
APPGSGKTTLLIKVCFETFSKNSLITLAKSRSEELAKVNRIVLDESDTLPQDRR
LITLDSYLMNRGLTCVLYLIDBCFVAGAAVACIEFTKCSALIFGDSRIRGRG
SELDITAVLSDNRFVDESRYGVSEYKCDVCMAMLSFPYKTVATMLVHAQGSMT
QVREIASEVDYSEFVYLTMLQSEKDDLKSGFKGRSRSEKPTVLVHAQGSMT
RKXVNLTRKQEBDFRSENHITVALSRHVESLYGVLSKRDALIAKIVAKOIVD
ARYRTPSPGSGSTLDSVSNPSTSDRSKCSASAPYVINSFLESVPVPGTTSYDRGVS
BEMKTOVPEBGADNVYRDSAPVNSKTDHENYYSBSRYNPTVVERBSGQBPQGA
MAMWLBERSFDEKAVANRSDVLAITKGRYTMKSRREPQDALSDDIQRNLNLEB
ITTRKLWKRDAKYLDSCLVHPAPQVMPFRKAVNALFSPCEPDEFNORPLTNGS
NIVFTEMINSTASIAKEMLGSEHYVNGEIDPFSEKSDQAFIKSFERTIYSAFGF
DEDLIDWMOGEYSNATITLDQSEFSDVQRKSGSNWINGISFETLILMEFYTN
RFKALFVSGDLSLIFESPRIENSADACTELGETKFLPSVYFCSKFFVWNGDVF
FVDPYKLVLYLQASKDEVDDELPEVFTSRDLTKDLYDERVIELLTHLVSKGYE
SGDLYALCAHCRSNFSSFKLIPYKVMVHVGKLVAKKFRANCRERAPDPAFG
ERTFLITTKLETYV"
exon
108..7983
/gene="ORF1a/b"
/number=1
7983..8045
/gene="ORF1a/b"
/number=1
7983..8045
/note="+1 ribosomal frameshift signal"
intron
7984..8044
/gene="ORF1a/b"
/number=1
8045..9393
/gene="ORF1a/b"
/number=2
9444..9608
/note="unnamed protein product; putative membrane-binding
protein"
CDS
/codon_start=1
/protein_id="CA51864.1"
/db_xref="GI:405627"
/db_xref="SPTREMBL:008542"
/translation="MDCVLRSYLLAFGLICFLPCLVVFVIFVYKQIILFRTAQS
N
EARNHSTVIV"
9608..11404
/codon_start=1
/product="heat shock 70-related protein"
/protein_id="CA51865.1"

```

```

/db_xref="GI:405628"
/db_xref="SWIS-PROT:P37092"
/translation="MVVFGIDFGTTPSSVCAVYGEELVFKORDSAVIPTVFLHSDT
OEVAFGVDAVLSNDLSVGFYRDLKRMIGCDSEERYDYLEKLPKHYTELKTAOS
SKSTVLKDCVSGPONATLPGLIATFVALISTASBAKCOCTGIVISVAPNAYCLO
RSPTESCYNLSGRCVYVNMNEPBAALASCSRISKATSPVLYDFDGGGTFDVSVAL
MNTFVYASGDMNLGRIIDKAFVHLYNKKQIPNNYKIDISPLKESKVSPLNE
PVYSEQGRVDVIVNSELAEVAPVEETIKIVERYKTCSSKRLERNAVAKULVY
GSSVYPLGLISRLSIIPVDECLVLDAAVAAGCALYSACLNRDSPLLVDCAHN
LSISKRCEIVCPAGSGPIPTGVATVMTQSNMNAVYSAALFEGDPFKCRLNIRIF
FGVVLVAGNVEYVTSARVPLTELINVSVGTISFSLVPTGVKILIGMAVDDSY
OLGERVADILKNSDKVLIHALITYOPORCKLTQDQXALFLKLTADYREARKFS
SYDDAVYINLHGR1IPKRLNGSVEKLAV"
11310..12971
/note="unnamed protein product; putative heat shock
70-related protein"
CDS
/codon_start=1
/protein_id="CA51866.1"
/db_xref="GI:405629"
/db_xref="GOA:008541"
/db_xref="SPTREMBL:008541"
/translation="MTTRESPTANVYWGIFRRFGCGEMKLMSEPAVSVPYSSD
FRSDGVILSRKTPGSGTGESFTRRESLILTPKTYEYKLGVMALNEMNRSD
YNSEEVIVVKTGCKRFNIOQVTERVKINGVAPSLVCHQWSJNSGCELINFDK
TKNPFVSLIFGKOLASTDEBAVSSYLDYLSHCNLNLYTCNLSVSGKSLYDEFLK
HVIDYENSDLEVRSPDNPLVAGIYDMCFEYNLTLSYLNIESFDFSLYPL
SEVYNSMMEPADVRLDFELDAEELIKVPTINMGDSFFLYNKLRYLESYFEDPSN
ELIKVVDLSLTNDNPELKLACQWVFGHCYGVFRACRYKVRDAEYLLPALGFFV
INMGVAFPEEELQKMPISVYRRCGSLSHEAFVFRGVPFPIRLVVPVYS
VLNVDYRHRKRGTLTODELTILSNIEPPIVAVMCCREVALDARRAQREKFGQWKG
TKNEISPHASSRYVKKKNDSLNITMKVAGASQRANPLRKR"
12919..13889
/codon_start=1
/product="coat protein homolog"
/protein_id="CA51867.1"
/db_xref="GI:405630"
/db_xref="GOA:008538"
/db_xref="SPTREMBL:008538"
/translation="MLAPEKRGDLHFTNTSDAMETFENSYDLAEVSNPNKLRK
ETBELGVYIERKSEVLTIDEDPVHGLPALIRANITTSYKYNVAVAYETTGKKG
FLVADWAFPLIECKKREKPPVPTFCATDEDAIVARSLPCLPRLRTIGRGIP
SGYEFLADQSLITVCLNDEHRAIVLQASRAAIDRAVSSVDGKIVSLFLDGRS"
13640..14254
/codon_start=1
/product="coat protein"
/protein_id="CA51868.1"
/db_xref="GI:405631"
/db_xref="GOA:008531"
/db_xref="SPTREMBL:008531"
/translation="MGSABPISAIATFENVSIADQTLGHGDCDKLRNFEELKLG
VPELDNLGIALGLCYSCATIGTSNKVNVPTSPFIASGGKELVLTGHELNSTLGS
OKLLEGKPNLRCFCRTPFCMDYSLKRGKGLPIPIARNRGLPDEHDYLAADPIS
STETLDQQRLLIAREMATHTEFSSESPTSLKQGRGLGGR"
14251..14793
/note="unnamed protein product; orf"
CDS
/codon_start=1
/protein_id="CA51869.1"
/db_xref="GI:405632"
/db_xref="SPTREMBL:008544"
/translation="MTSVEVLAQTKPLFVLLKGFVYVVAFFETEESSEAEPLVY
LHDELINIRKGIASVYDFMGCMTLRPSSVYTRVSGSESSDFSLPGSGKTFGS
KVLNRKVTFTFENGVOLVFGMYGLBQRCSVDIWEANFVGAHGTLLYCLNCELDK
SGGELITLFSKVEVLKRW"
14768..15301
/note="unnamed protein product; orf"
CDS
/codon_start=1
/protein_id="CA51870.1"
/db_xref="GI:405633"
/db_xref="SPTREMBL:008545"
/translation="MKFELKDGTSRLSRSSEILRKYVELGNSQOSEISCVDFEN
ELASPNLIVTVHERMEQHPOSSKLVPSIGSMLEIRAPLKVRYVTMHEKTA
SDTLNAPLEBYCKITGLAREDALEBQRKRYKSVYVLPHEBELKPEVTMMPSTYELK

```

3'UTR
INTSLRVISQILGWAI"
15299..15480

Query Match 10.7%; Score 63.8; DB 14; Length 15480;
Best Local Similarity 48.1%; Pred. No. 8.4e-09;
Matches 213; Conservative 0; Mismatches 227; Indels 3; Gaps 1;

QY 95 TTGAAAAATGTTGGTGCAGAAAGGGGCTCCTAACGAGGTAAGAAAGTGTGTTGGTGC 154
DB 13746 TCGAAGAGTGTGTAATTAAGAGGTTCCGAAAGATTAACCTCGAATCGCGTTAGAC 13805
QY 155 TACTCCTTTAGCACTCGCGGCAAGACAGTCTCCCAAGTTCA---GGCGCAANTT 211
DB 13806 TTGTGTTGATTCCTGTGTGTGATGAGTACCTTCAACAAAGTTAAGTCCACCGAGT 13865
QY 212 CAGAGCTATATTTTCAATAGTTTCGAGAGAGATGTGTATGATTAACAGAGGTGACC 271
DB 13866 CTACCTTCATCAAGCTTGTTGTGTGTGGGAAGAACTGACCTCACTCAAGGTGAAT 13925
QY 272 TTAAGAAAGTACGACGAGTGTGGCCTTCACAGTTCAATTAATTAAGTGAACGT 331
DB 13926 TGAATTCCTTTCTGGAGTCTCAAAAACCTTTGAGAGGAAACCTTAACAAATTCGAGT 13985
QY 332 TCGGTCGATCTTCACTAGAGGCTTACGTTGACTTTGTATCGCGTATPAGACAAATTAC 391
DB 13986 TCTGCCGATCTTTTCAGAAAGGACTACATATCTTTCGCAAGAAATACCGAGGAAATTAC 14045
QY 392 CCCAATCAAGCGCGCGGGAATTTGGGATTCGAGGTAAGTGTGTAAGTGTGAGTGCAG 451
DB 14046 CTCGCGATTCGAGAGTAAACCTGACGCTTACCGCGTGAAGATCACTACTTACCGCTG 14105
QY 452 ATTTCTGGGTAATTGCCGGAAGCTCTGTGAATTAACAGCAAGTAGAAGATGTTCCGA 511
DB 14106 ACTCATATCGACGTCGACGAGTACTGACTGACCAACAAAGTCTGCTGTATAGCGC 14165
QY 512 GTATGAGCGCTTAAAACTGAA 534
DB 14166 GCGAAAGCGCACTCACAGGAA 14188

RESULT 13
LOCUS A41914 783 bp DNA linear PAT 05-MAR-1997
DEFINITION Sequence 1 from Patent WO9428147.
ACCESSION A41914
VERSION A41914.1 GI:2297469

KEYWORDS
SOURCE Beethellows virus
ORGANISM Beethellows virus
REFERENCE Beethellows positive-strand viruses, no DNA stage;
Closteroviridae; Closterovirus.
1 (bases 1 to 783)
AUTHORS Bojse, K. and Brunsfeldt, J.
TITLE DNA SEQUENCE COMPRISING AT LEAST TWO COAT PROTEIN GENES
JOURNAL Patent: WO 9428147-A 1 08-DEC-1994;
SANDOZ AG (AT)
COMMENT Other publication AU 7122694 941220.
FEATURES
location/Qualifiers
source 1..783
/organism="Beethellows virus"
/mol_type="unassigned DNA"
/db_xref="taxon:12161"

ORIGIN

Query Match 10.6%; Score 63; DB 6; Length 783;
Best Local Similarity 46.6%; Pred. No. 9.7e-09;
Matches 228; Conservative 0; Mismatches 270; Indels 3; Gaps 1;

QY 26 TTAGCAACTGCTGATTAACCGACGCTCTAGCTTAATGATGTGCAGACAAAGCTTTAT 85
DB 126 TTGAAAGCTGAGTCTGACACCAAGCTGTTTCCAGCGTGAAGACTCGCAAAACTTAC 185
QY 86 CTGCTGAAGTTGAAAAATGTTGTGTGCAGAAAGGGGCTCTTACAGAGGTATAGAAAGTGG 145

DB 186 GGAAGAAATTCGAAAGAGTGTGAATTAAGAGGGGTTCCGAAGCAAACTCGTCTCG 245
QY 146 TGTTCGCTACTCTCTTAAAGCACTCGCGGCAAGAACAGTCTCTTAAGTTCAGCGG 205
DB 246 GGTAAAGCTTGTGTTGATTCCTGTGTGAGAGTAAAGTACTTAAATTAAGTGTGTC 305
QY 206 CAGATTCAGAGCTTATATTTTCAATAGTTTCGAGAGAGAAATG--TGTATGTAACAG 262
DB 306 AACGACGCTACTTCAATCAAAAGCTTCGTCGATGTGTGGAAGAAATGTTCTCATCTC 365
QY 263 AGGTGACCTTAAGAGTACTGACGAGTGTGCGCTCTCACTAGTTCACATTAATAC 322
DB 366 ACGGTAACCTAGGCTCTTTCTGAGCTCTGAGAAACCTTTGAGAGAAAGCTTAACAGT 425
QY 323 TTAGAACGTTGGTGTACTTCACTGAGGCTTACGTTGACTTTGTATCGCGTATPAGC 382
DB 426 TCGGTTGTTTCTGCGGCACTTTTCAGAAAGACTATATCTTCGCAAGAAATACCGAG 485
QY 383 ACAATTAACCCCACTCAACGCGCGGCGGAATTTGGGATTCAGAGTGAAGATTCGACT 442
DB 486 GAGAGCTGCTTCGATTCGATTAAGCCCAACGCTCACTGCTGAGAGTCACTACT 545
QY 443 TAGCTGCAAGATTTCTGGGTAATTGCCGGAAGCTCTCTGAATTAACAGCAAGTAGAAGA 502
DB 546 TAGCTGCTGATTCATATCGACATCAACAGAACTTACTGACTACCAACAAAGTGTCTGC 605
QY 503 TGTTCGCGAGTATGACGCTTAAAACTGA 533
DB 606 TGTTCGCGCGGAAAGCGCACTCACACAGA 636

RESULT 14
LOCUS BYVMBPA 5980 bp RNA linear VRL 01-OCT-1993
DEFINITION Beethellows virus (isolate German BYV-G) genes for putative
membrane-binding protein, heat shock 70-related protein, coat
protein homolog and coat protein.
X73475.1 GI:313689

ACCESSION X73475.1 GI:313689
VERSION X73475.1
KEYWORDS coat protein; heat shock 70-related protein; membrane-binding
protein.
SOURCE Beethellows virus
ORGANISM Beethellows virus
REFERENCE Beethellows positive-strand viruses, no DNA stage;
Closteroviridae; Closterovirus.
1

Agranovsky, A.A., Koonin, E.Y., Boyko, V.P., Mais, E., Lunina, N.A. and
Atabekov, J.G.
TITLE Beethellows closterovirus: complete genome structure and
identification of a leader papain-like thiol protease
unpublished
2 (bases 1 to 5980)

REFERENCE Mais, E.
AUTHORS Direct Submission
TITLE Submitted (18-JUN-1993) E. Mais, Inst. f. Biochemie und
Pflanzenvirologie, Biologische Bundesanstalt, Institut fuer
Biochemie, Messeweg 11/12 D-3300 Braunschweig, FRG
JOURNAL location/Qualifiers

FEATURES
source 1..5980
/organism="Beethellows virus"
/mol_type="genomic RNA"
/isolate="German (BYV-G)"
/db_xref="taxon:12161"
/clone="pBYV-D22, pBYV-D31"
71..235
/note="unassigned protein product; putative membrane-binding
protein"
/codon_start=1
/protein_id="CA51857.1"
/db_xref="GI:313690"
/db_xref="SPTREMBL:O65893"
/translation="MDCVLSRYLLAFGLICFLFCLVFIWFYKQILFRNPDSN"

CDS

CDS
EARNSTVTV"
235...2031
/codon_start=1
/product="heat shock 70-related protein"
/protein_id="CA51858.1"
/db_xref="GI:313691"
/db_xref="GOA:Q65894"
/db_xref="SPTREMBL:Q65894"
/translation="MYVFGDDEGTTPSSCAVYGEELVLFKORDSAVIFPVEFLSDT
QVAPRYDAVLSNDSSVGRGFRDLKRWIGDDEMYGDLKPKHYTELKXAOS
SKSTKLDGSGVFPQATLPGLITATFYALLISTABAKCCCTGTCVSPANVACIO
RSFTESCVNLGSGPCVYVWNEPSAALISACRSIKGATSPVLYDGGTFVSVISAL
NNTFVVSAGSGMGLGDRIDKAFVHLKKQAPVNYKIDISFLKESLSKVSFLNF
PVSEONVADVLYVNSVLAELVAPVEERTIKVNEVYKRGSMLEPSVAKILNV
GGSYLPGILSRSLSVFEBCVLDPDAAVAGCALYACLRDPSMLLVDCAAH
LSISRYCESIVCPVAGSPIPFTGRTVMNMGASAVSALFEDFPKCLNRIIF
SGVVALGVGVSGSNRTVPLTLEINVSAGITTFSLVGPVTKLVGNMAAYDSY
QLEERVADLHKNSDKVITHALITYKPPQKLTLDSDALFLKLSADYRREARKFS
SYDDAVNSSELLGRVLPKLRGSRVEKLDV"
1937...3598
/note="unnamed protein product; orf4"
/codon_start=1
/protein_id="CA51859.1"
/db_xref="GI:313692"
/db_xref="SPTREMBL:Q65895"
/translation="MTFRSTPANYTWGELFRFPFGREKMLMSFAVRRPYARD
FRPSNGVLSRKIFSDTGSFVREFSLTLPKTYEVCCLGVANEQALSNGRLSD
YVNSLAVDVKTVCKFNQIVTEFVKLNGMLAPSLVHWSLNSGCLINPKD
TKEFVSLIFPKQVIEBTDIAIVSSLYDLSHCLNLYETCNLSNSGKALYDEFLK
YVLYLETSLDEHSPSDNPDLVAGVLYDMCFEYNTLSTYKNIISFDFCLYPL
SEIFSNMWBOPADPVLPELDTAELLIKPTINTLSTFLYKRLRYESYEEDSN
ELIKVQVDSLTRDNPETLKAQWVGHCHYGVFRACQKRYADVAPYCPALGPT
LNNSGVEFDELQKMPVSVRKRGSGSHAFSTFRFGVGPFPPEYRRLNVPYYS
LYNVYRVHVRKGLQDELTLISNIEFDVAECCEVALQRRQOREKPEQMGK
VKNVESPHARSSIRVKSNSLNLIMKDVGRQRLNPLNRKH"
3546...4196
/codon_start=1
/product="coat protein homolog"
/protein_id="CA51860.1"
/db_xref="GI:313693"
/db_xref="GOA:Q65896"
/db_xref="SPTREMBL:Q65896"
/translation="MLASGRGLIHFTENTRDAMETFFNSYDLAEYSEVNPKNLRK
ETBELGIVIRERFSELYITDEPVEHLAFLIRANITTSVTNVVAGAEYVIGKK
FLVYDAVPEPLIKECMKENKPNPVEPCATFEDAYIVLARSLPKLIRNTGKGLP
SGIFLGADFLIATVCLNDHEKAIYVQASRAIDRAVSSVQKIVSLFLGRLS"
4267...4881
/codon_start=1
/product="coat protein"
/protein_id="CA51861.1"
/db_xref="GI:313694"
/db_xref="GOA:Q65897"
/db_xref="SPTREMBL:Q65897"
/translation="WGSAPFLSATATPENVSADQTLHEDCDKLRNPEGLKIKG
VPEBKGLALGLCLVSCATIGTNSKVSQVPSITFKSPGSGKELVTHSELFSIDS
QKLEGPKNLRCFCRTFKDYISFAKEYGRPLRARRHGLPMDHYLADP1ST
STELTDQGRGLIARENATHTERSSESPTYSKQLORGLATGS"
4878...5411
/note="unnamed protein product; orf7"
/codon_start=1
/protein_id="CA51862.1"
/db_xref="GI:313695"
/db_xref="SPTREMBL:Q65897"
/translation="MIGVELACTRPFRRVLLKGVFYIVALETEEPESEALPLVY
LHDELVNKKGISSYIDPKSCMRLLKSSVSATVSSGKSEDFSSSGSKTES
RIINRKYTFENGOLVFGMYGRDRCVSESEVLFENVVGAAHCGTLPYCLNCELDK
SGDLEITFSKNEVFL"
5395...5928
/note="unnamed protein product; orf8"
/codon_start=1
/protein_id="CA51863.1"
/db_xref="GI:313696"
/db_xref="SPTREMBL:Q65898"
/db_xref="SPTREMBL:Q65898"

ORIGIN
/translation="MKFFENDGTSALRSBSLLRKYELGNTNSQSEVSECTNEFN
ELARFNFLVTVHEHREMEHGRQSSSELRAPSLGLMLKEIRFLKRVVTPMKETA
SETLNAFLDEYCRJAGLITREDALREKWRKVSIVLPFHSHLELKEFVENTENFSTELLK
LNLRLRVSSQIIIGIAV"

Query Match	10.6%	Score 63	DB 14	Length 5980
Best Local Similarity	46.6%	Pred. No. 1.3e-08		
Matches 238	Conservative 0	Mismatches 270	Indels 3	Gaps 1
26	TTAGAACCTGGTATTAACGACGCTCTAGCTTAATGCTGTCACAGAAAGCTTTAT	85		
4304	TTGAAGAGTACGCTCGACAGCAACCGTGTTCACGGTGAAGACTGGACAACTAC	4263		
86	CTGCTGAAGTTGAAAATGTTGTCGCAAAAGGCGCTCTTACGAGGATTAAGATG	145		
4364	GGAAGAAATTTGAAAGAGTGTAAATTAAGAAAGGGGTTCGGAAGACAACTCGTTCG	4423		
146	TGTTGGTCTACTCTCTTACGACCTCGGGCAGAACACGCTCTTACGAGTTCAGCGCG	205		
4424	CGTTAGACCTTTGTTTGTATTCCTGTGCGACATAGTACTCTTAATTAAGTTAGTCC	4483		
206	CAGATTCAAGCTATATATTTCAATAGTTTGGAGAGAGATG---TGGTACTAACG	262		
4484	AACGACGCTACTTTCATCAACGCTTCTGCTGGTGGAGAGAAATTTGTTCTCACTC	4543		
263	AGGTGACCTTAAGAGGACTTCACGCGGTGTGGCCCTTCACATAGTTCACTATAAC	322		
4544	ACGGTGAAGTGAAGCTTTTCTGACCTCTCAAAACTTTTGAAGAAAGCTTAACAGT	4603		
323	TTAGAACCTGGTGTGCTACTTCACTAGAGCTTACGTTAGTCTTTGATCGGATTAAGC	382		
4604	TCCGTTGTTTTCGCGGACCTTTTGAAGAGACTACATATCTTCGCGAAGAAATACGAG	4663		
383	ACAAATTAACCCCACTCAACCGCGGGGGAATGGGATTCGACGTAAATTCGTA	442		
4664	GAAAGCTGCTCCGATTCCTAGAGCAACGCTCACGGTCTACGCTGAAGATCACTACT	4723		
443	TAGCTGACGATTTTGTGGTACTTTCGCCGAGCTCTGTAATTAACGAAAGTGAAGA	502		
4724	TAGCTGCTGATTTTCAATTCGACATCAAGAACTACTGACCTAACAGAGTGTCTGC	4783		
503	TGTTGCGAGATATGACCTCTTAATAACTGA	533		
4784	TGTTGGCGGCGGAAACCGCACTCACACAGA	4814		

RESULT 15
AX655393/C
LOCUS AX655393 2000 bp DNA linear PAT 22-MAR-2003
DEFINITION Sequence 5263 from Patent WO03000898.
ACCESSION AX655393
VERSION AX655393.1 GI:29158207
KEYWORDS
SOURCE
ORGANISM
Oryza sativa
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzoideae; Oryza.

REFERENCE
AUTHORS Chang, H.S., Chen, W., Cooper, B., Glazebrook, J., Goff, S.A., Hsu, Y.M.,
Katagiri, F., Quan, S., Tao, Y., Whitman, S., Xie, Z., Zhu, T. and Zou, G.
TITLE Plant genes involved in defense against pathogens
JOURNAL Patent: WO 03000898-A 5263 03-JAN-2003;
Syngenta Participations AG (CH)
FEATURES
source
1..2000
/organism="Oryza sativa"
/mol_type="unassigned DNA"
/db_xref="taxon:4530"

ORIGIN
Query Match 6.3%, Score 37.8; DB 6; Length 2000;

Best Local Similarity 9.4%; Pred. No. 1.7;
Matches 48; Conservative 231; Mismatches 228; Indels 2; Gaps 1;

```
QY 16 GACAGAGAACTTGAAGCACTGTGATTAACGACGCTCTAGTCTAAATGATGTCGACAAG 75
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1009 SWSWTYAMMKRYKMTAISTWKMYAKMRAYAMSRKTIWCTGGRNATTCGTAKM 950
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 76 AAGCTTTTATCTGCTGAAGTTGAAAAATGTTGTGACAGAAAAGGGCTCTTAACAGAGGT 135
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 949 AAGRMNRBAMCWCYCCMMKWKMTSCMMWKYRMTSCWYTMWGAARAYAYAMRRRBMTY 890
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 136 ATGAAAGTGTGCTTCGCTTACTCTTACGACCTCGCGCAAGAACAGCTCTCTCTAG 195
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 889 KMSWRMYTMTMTAMTMTCMCAKMYATGNATMMMYTMTYCYAMTCAKCKYKA 830
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 196 GTTCAGCGCGCAGATTCAGAGCTTATATTTTCAATATGTTTCGAGAGAGAGATGTGTA 255
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 829 MTKMWTITWACAMRALTISRMBAMAGMWRKRYKMRBAYWWRWRCWKAGWABMMKSRYBMW 770
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 256 GTAACAGAGGGTACCTTAAGAAAGTACTGACGGGTGTGCGCTCTCACTAGTTCAGT 315
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 769 KKYATRYMWMAMTMMWSWRKSYRMSGMRMRMSAMRYCSRMKCAKTKYASSARMT 710
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 316 AATAAAGTGAACGTTGCGTCTGACTTCTGAGAGCT--TACGTTGACTTTGTATCG 373
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 709 KRAQSYRYRRRWYKWKGTIRYRYSRSMTRABMSKRKRWAGASMSQWYMRGAS 650
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 374 CGTATAAGCACAATTACCCCACTCAAGCGCGCGGGAATTGGGATTCAGCTGAAG 433
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 649 MWYSKYSKSAKCKKTRMTSYMTSGMYSSYKMSMTSKMSYVGRKTCMTYTSMKG 590
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 434 ATTGCTACTAGCTGCAGATTTTCTGGTACTGCCGAGAGCTCTGGAATTAAGCAAA 493
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 589 STRSSKMRSGMSRMTYRMRKRRKRYMKKCTWRKWCYRWGYMTTSSRPM 530
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 494 GTAGAAGATGTTGCGAGATATGTACGCT 522
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 529 YTGRYKARYTSKRRYMWYKRYKWCYYY 501
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: April 7, 2004, 00:33:02
Job time : 2193 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 19:32:38 ; Search time 304 Seconds
(without alignments)
8342.678 Million cell updates/sec

Title: US-09-613-486-14

Sequence: 1 atgaggttcgctccgacag.....gtagaagcgaagtatgtaa 597

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*

1: geneseqn1980s:*\n2: geneseqn1990s:*\n3: geneseqn2000s:*\n4: geneseqn2001as:*\n5: geneseqn2001bs:*\n6: geneseqn2002s:*\n7: geneseqn2003as:*\n8: geneseqn2003bs:*\n9: geneseqn2003cs:*\n10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	597	100.0	597	2	AAV08870 Grapevine
2	597	100.0	15500	2	AAV08874 Grapevine
3	63.8	10.7	688	2	AAQ87853 Sugar bee
4	63	10.6	783	2	AAI42975 BYV coat
5	37.8	6.3	2000	7	ADA71938 Rice gene
6	36.8	6.2	1332	6	ABN67902 Streptococ
7	36.8	6.2	11000	6	ABN71527_13
8	36.6	6.1	1605	7	ACA38387 Prokaryot
9	36.6	6.1	11000	4	AAI99682_18
10	36.6	6.1	11000	4	AAI99683_18
11	35.6	6.0	593	7	ABX74308 Human rep
12	35.6	5.7	6756	4	ABL04577 Human tes
13	33.8	5.7	6756	4	ABL97500 Human tes
14	33.8	5.6	360	4	AAI82806 Human pol
15	32.6	5.5	972	4	AAI14902 Human tra
16	32.6	5.5	1074	4	AAI14901 Human tra
17	32.6	5.5	1170	4	AAI14893 Human tra
18	32.6	5.5	1245	4	AAI14900 Human tra
19	32.6	5.5	1272	4	AAI14892 Human tra
20	32.6	5.5	1443	4	AAI14891 Human tra
21	32.6	5.5	2000	7	ADA71938 Rice gene
22	31.6	5.3	8327	6	ABK39669 Human che
23	31.2	5.2	384	6	ABK73751 Bacillus

C 24	31.2	5.2	5920	4	ABL01884
C 25	31	5.2	641	5	AA876083
C 26	31	5.2	2304	6	AA888852
C 27	31	5.2	3955	2	AAV15181
C 28	31	5.2	4315	2	AAV63754
C 29	31	5.2	4483	2	AAI40091
C 30	31	5.2	4483	3	AA895047
C 31	31	5.2	4702	6	AA895047
C 32	31	5.2	4749	9	ADD00948
C 33	31	5.2	4974	9	ADD00955
C 34	31	5.2	14446	3	ABN68600
C 35	30.8	5.2	1134	6	ABN71527_21
C 36	30.8	5.2	5561	6	ABN71527_21
C 37	30.4	5.1	454	6	ABV86742
C 38	30.4	5.1	498	6	ABX55372
C 39	30.4	5.1	591	5	ABV55404
C 40	30.4	5.1	1010	7	ABZ35965
C 41	30.4	5.1	1492	6	AA835505
C 42	30.2	5.1	191	4	AA850560
C 43	30.2	5.1	191	7	ACA17841
C 44	30.2	5.1	460	4	ABA74297
C 45	30.2	5.1	460	4	AAI54759

ALIGNMENTS

RESULT 1
AAV08870 ID AAV08870 standard; cDNA; 597 BP.
XX AC AAV08870;
XX XX
DT 27-AUG-2003 (revised)
DT 29-MAR-1999 (first entry)
XX XX
DE Grapevine leafroll virus type 2 coat protein ORF6 product.
XX GRLV-2; closterovirus; grape; tobacco; transgenic plant;
XX disease resistance; virus resistance; beet yellows virus; tristeza virus;
KW coat protein; ss.
XX OS Grapevine leafroll virus.
XX XX
XX MO9853055-AI.
XX PD 26-NOV-1998.
XX PF 20-MAY-1998; 98WO-US010313.
XX PR 20-MAY-1997; 97US-0047194P.
XX PA (CORR) CORNELL RES FOUND INC.
XX PI Zhu H, Ling K, Gomesalves D;
XX MPI; 1999-045307/04.
XX P-PSDB; AAV73482.
XX DR
XX PT Grapevine leafroll virus (type 2) proteins and polypeptides - and
XX encoding DNA, useful e.g. to impart grapevine leafroll resistance to
XX grape and tobacco plants and detect grapevine leafroll virus.
XX Claim 38; Page 44; 151P; English.
XX
XX This is the nucleotide sequence of open reading frame ORF6 of grapevine
XX leafroll virus type 2 (GRLV-2) RNA (see AAV08874). It codes for a 22 kDa
XX coat protein (see AAV73482). The GRLV-2 genome includes 9 open reading
XX frames (see AAV08864-72) for a polypeptide, an RNA-dependent RNA
XX polymerase, heat shock proteins, coat proteins and proteins of unknown
XX function (see AAV73476-84). These can be used to produce antibodies
XX useful for detecting GRLV in samples e.g. by ELISA (claimed). The
XX nucleic acid molecules can be used to produce probes and primers for such

CC detection, and to transform host cells (especially *Agrobacterium* *vitis*,
 CC *Agrobacterium tumefaciens*, grape, citrus, beet or tobacco cells) and
 CC produce transgenic plants (claimed). They can be used to impart GRLAV-2
 CC resistance to *Vitis* ssp. or rootstock cultivars or *Nicotiana* cultivars
 CC (claimed). Because extensive similarity exists between hsp70-related
 CC sequence regions of GRLAV-2 and other closteroviruses, the DNA may also
 CC be used to impart beet yellows virus resistance to beet cultivars or
 CC *tristeza* virus resistance to citrus ssp. cultivars/rootstock cultivars
 CC (claimed). (Updated on 27-AUG-2003 to correct OS field.)

XX Sequence 597 BP; 165 A; 121 C; 157 G; 154 T; 0 U; 0 Other;

Query Match 100.0%; Score 597; DB 2; Length 597;
 Best Local Similarity 100.0%; Pred. No. 6,4e-193;
 Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGATTGATGTCGACGACCACTTACGCACTTGATGATACCGACGCTTACTGTA 60
 Db 1 ATGAGATTGATGTCGACGACCACTTACGCACTTGATGATACCGACGCTTACTGTA 60
 QY 61 AATGATGTCGACGACCACTTATCTGCTGATGATGAAAAATGTTGTGCAAAAAGG 120
 Db 61 AATGATGTCGACGACCACTTATCTGCTGATGATGAAAAATGTTGTGCAAAAAGG 120
 QY 121 GCTCTTAACGAGGGTATGAAAGTGTGCTTACTCTTTCGCACTCGCGCAAGA 180
 Db 121 GCTCTTAACGAGGGTATGAAAGTGTGCTTACTCTTTCGCACTCGCGCAAGA 180
 QY 181 ACCACGTCCTTAAGCTTACGCGCGGAGATTACAGCTTATTTCAATAGTTGGA 240
 Db 181 ACCACGTCCTTAAGCTTACGCGCGGAGATTACAGCTTATTTCAATAGTTGGA 240
 QY 241 GAGAGGATGTGTAGTATACAGAGGTCACCTTAAGAGGTAATCGAGGTCGCT 300
 Db 241 GAGAGGATGTGTAGTATACAGAGGTCACCTTAAGAGGTAATCGAGGTCGCT 300
 QY 301 CTCACGATGTCATTAATTAACCTTAAAGCTTCTGCTGCTTACGAGGCTAGCT 360
 Db 301 CTCACGATGTCATTAATTAACCTTAAAGCTTCTGCTGCTTACGAGGCTAGCT 360
 QY 361 GACTTTTGTATCGGTATATACCAAAATTAATCCCAACTCAACGCGCGGGAATGGGG 420
 Db 361 GACTTTTGTATCGGTATATACCAAAATTAATCCCAACTCAACGCGCGGGAATGGGG 420
 QY 421 ATTCAGCTGAAGATTCGTAATCTTACGCTGCAATTTTCGGGTACTTCCCGAAGCTCT 480
 Db 421 ATTCAGCTGAAGATTCGTAATCTTACGCTGCAATTTTCGGGTACTTCCCGAAGCTCT 480
 QY 481 GATTACAGCAAGATGAGAGATGTTCCGAGTATGTACGCTCTAAATAATGAAAGTGA 540
 Db 481 GATTACAGCAAGATGAGAGATGTTCCGAGTATGTACGCTCTAAATAATGAAAGTGA 540
 QY 541 GTGTAAATTAACCAAGTACGATTCGCTACAGCTAGTAAAGAGGAAGTATATA 597
 Db 541 GTGTAAATTAACCAAGTACGATTCGCTACAGCTAGTAAAGAGGAAGTATATA 597

RESULT 2

ID AAV08874 standard; cDNA; 15500 BP.

AAV08874;

27-AUG-2003 (revised)
 29-MAR-1999 (first entry)

XX Grapevine leafroll virus type 2 (GRLAV-2) genome.

XX GRLAV-2; closterovirus; grape; tobacco; transgenic plant;
 XX disease resistance; virus resistance; beet yellows virus;
 XX protease; methyltransferase; helicase; heat shock protein; coat protein;
 XX RNA polymerase; se.

OS Grapevine leafroll virus.

XX Key Location/Qualifiers
 XX 4..7923
 FT /tag= a
 FT /product= "polypeptide (protease, methyltransferase,
 FT helicase)"
 FT /note= "Claim 26"
 FT 7922..9301
 FT /tag= b
 FT /product= "RNA-dependent RNA polymerase"
 FT /note= "Claim 29"
 FT 9305..9535
 FT /tag= c
 FT /product= "hydrophobic protein"
 FT /note= "Claim 43"
 FT 9551..11350
 FT /tag= d
 FT /product= "heat shock 70 protein"
 FT /note= "Claim 32"
 FT 11277..12932
 FT /tag= e
 FT /product= "heat shock 90 protein"
 FT /note= "Claim 35"
 FT 12844..13515
 FT /tag= f
 FT /product= "diverged coat protein"
 FT /note= "Claim 41"
 FT 13584..14180
 FT /tag= g
 FT /product= "coat protein"
 FT /note= "Claim 38"
 FT 14180..14665
 FT /tag= h
 FT /product= "undefined"
 FT /note= "Claim 45"
 FT 14667..15284
 FT /tag= i
 FT /product= "undefined"
 FT /note= "Claim 47"
 FT 15285..15500
 FT /tag= j
 FT /note= "Claim 48"
 PN WO9853055-A1.
 XX 26-NOV-1998.
 XX 20-MAY-1998; 98WO-US010313.
 XX 20-MAY-1997; 97US-0047194P.
 XX (CORR) CORNELL RES FOUND INC.
 PA Zhu H, Ling K, Goncalves D;
 XX WPI; 1999-045307/04.
 DR P-BSDH; AAW73476, AAW73477, AAW73478, AAW73479, AAW73480, AAW73481,
 DR AAW73482, AAW73483, AAW73484.
 XX Grapevine leafroll virus (type 2) proteins and polypeptides - and
 PT encoding DNA, useful e.g. to impart grapevine leafroll resistance to
 PT grape and tobacco plants and detect grapevine leafroll virus.
 XX Disclosure; Page 12-20; 15pp; English.
 XX A total of 15,500 bp of the RNA genome of grapevine leafroll virus type 2
 CC (GRLAV-2) has been sequenced and cloned from GRLAV-2 isolated from
 CC infected *Vitis vinifera* cv. Pinot Noir. About 85% of the total RNA genome
 CC was revealed from 2 different clones. The sequence in the coat protein
 CC gene region was determined and confirmed from several overlapping clones.
 CC The genomic organization of GRLAV-2 includes 9 open reading frames (see
 CC also AAV08864-72) encoding a polypeptide (see AAW73476) having papain-

QY 512 GTATGTACGCTCTTAAATCTGAA 534
 Db 600 GCGAAAAGCCTACACGCGAA 622

RESULT 4

AAAT42975 standard; cDNA; 783 BP.

AAAT42975;

16-OCT-2003 (revised)

14-JAN-1997 (first entry)

BYV coat protein cDNA.

Tobacco mosaic virus; coat protein; beet yellow virus; BYV;

beet western yellow virus; BWYV; beet necrotic yellow vein virus; BNYYV;

luteovirus; closterovirus; sugar beet; fruit; mango; apple; pear; banana;

field crop; sunflower; wheat; barley; maize; vegetable; potato; carrot;

cabbage; onion; ss.

Beet yellows virus.

Key Location/Qualifiers

sig_peptide 13..78

/tag= a

/note= "Omega sequence from TMV"

89..703

/tag= b

/product= "BYV_coat_protein"

WO9428147-A1.

08-DEC-1994.

01-JUN-1994; 94WO-EF001786.

02-JUN-1993; 93GB-00011332.

(SANO) SANDOZ LTD.

(SANO) SANDOZ PATENT GMBH.

(SANO) SANDOZ-ERFINDUNGEN VERM GRS MBH.

Bojsen K, Brunstedt J;

WPI; 1995-022818/03.

Recombinant DNA sequence comprising at least two coat protein genes
 used to combat viruses in e.g. sugar beet.

Disclosure: Page 11-12; 23pp; English.

The sequences given in AA142975-77 represent cDNA sequences encoding coat
 proteins from beet yellow virus (BYV), beet western yellow virus (BWYV)
 and beet necrotic yellow vein virus (BNYYV) respectively. These sequences
 are used in the recombinant DNA sequence of the invention. This comprises
 nucleotide sequences encoding the coat proteins of at least one
 luteovirus and at least one closterovirus. This recombinant DNA may be
 used to transform sugar beet for combating viral infection. Other plants
 may also be transformed e.g. fruit such as mangoes, apples, pears,
 CC bananas, and field crops such as sunflower, wheat, barley, maize, and
 CC vegetables such as potatoes, carrots, cabbage and onion. (Updated on 16-
 CC OCT-2003 to standardise OS field)

Sequence 783 BP; 223 A; 178 C; 171 G; 211 T; 0 U; 0 Other;

Query Match 10.6%; Score 63; DB 2; Length 783;

Best Local Similarity 46.6%; Pred.No. 1.6e-10;

Matches 238; Conservative 0; Mismatches 270; Indels 3; Gaps 1;

QY 26 TTAGCAACTGGTGAATACCGAGCCTTAGTCTAATGATGTGCACAAGAGCTTTAT 85

Db 126 TTGAAAACGTGAGTCTCGCAGACCAACGTGTTTGACGCGTAGAGACTGCGACAACTAC 185
 QY 86 CTGCTGAAGTTGAAAAAATGTTGTGCGAAGAGGGGCTCTTAAGAGGCTATAAGTGG 145
 Db 186 GGAGAAATTTGGAAGAGTGTGAAATTTGAAATTTGAAAGGGGTTCCGAAAGACAAATCGCTTCG 245
 QY 146 TGTTCGTCCTACTCTTTACGCACTCGCGGAGAACCCGCTCTTAAGGTTCAAGCGCG 205
 Db 246 CGTTGAACTTTGTTTGTATTCCTGTGACAGATAGTACTTCTTAATTAAGTTAGTTC 305
 QY 206 CAGATTCAAGCTTATTTTCAATTAAGTTTCGAGAGAGGAAG--TGTAGTAAAG 262
 Db 306 AACGACGTCCTACTTTCATCAAAAGCTTGCTGGTGGAGAGAAATGTTCTCACTC 365
 QY 263 AAGTGACCTTAAGAAAGTACTCGACGGGTGCGCCTCTCACTAGTTCACTATAAAC 322
 Db 366 ACGTGAACTAGAGCTTTCTGAGCTCTCAAAACCTTTGAAAGAAAGCCTAACAAAGT 425
 QY 323 TTAGAACGTTGGGTGTACTTTCAGTGAAGCTTACCTTTGTATTCGGCTATAAGC 382
 Db 426 TCGGTTGTTCTGCCGCACTTTTCAGAAAGACTACATATCTTTCGCGAAGGAATACCGAG 485
 QY 383 ACAATTTACCCCACTCAACGCGCGGGAATTGGGATTCGAGCTGAAGATTGTA 442
 Db 486 GAAAGCTGCTTCGATTGCTAGAGCCAAACGTCAGAGTCTACCTGTAAATCACTACT 545
 QY 443 TAGCTGCAATTTTCTGGTACTTCCCGAAGCTCTTGAATTACAGCAAGTAGAAGA 502
 Db 546 TAGCTGATTTTCAATTCGACATCAAGAACTTACGACCTACACAAAGTGTCTGC 605
 QY 503 TGTTCGCGATGATGATCGCTCAAAACTGA 533
 Db 606 TGTTCGCGCGCAAAAGCTCACTACACAGA 636

RESULT 5

ADA71938/c

ID ADA71938 standard; DNA; 2000 BP.

ADA71938;

20-NOV-2003 (first entry)

Rice gene, SEQ ID 5263.

Plant; bacterial infection; fungal infection; viral infection; rice;

gene; ds.

Oryza sativa.

WO2003000898-A1.

03-JAN-2003.

22-JUN-2001; 2001WO-IB001105.

22-JUN-2001; 2001WO-IB001105.

(SYGN) SYNGENTA PARTICIPATIONS AG.

Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;

Katagiri F, Qian S, Tao Y, Whitman S, Xie Z, Zhu T, Zou G;

WPI; 2003-175290/17.

Identifying at least one gene involved in plant resistance or response to
 PT pathogenic infection for conferring resistance or tolerance to a plant to
 PT bacterial, fungal or viral infection by determining or detecting plant
 FT gene expression.

Claim 27; SEQ ID NO 5263; 899pp; English.

XX

CC The present invention relates to a method (M1) for identifying genes
 CC involved in plant resistance or response to pathogenic infection. M1
 CC comprises identifying a gene whose expression is significantly altered in
 CC the incompatible interaction of plant gene expression relative to
 CC the incompatible interaction of plant gene expression relative to
 CC expression of the gene in an uninfected plant, in a mutant plant that
 CC does not express a gene associated with response to pathogenic infection,
 CC or in a corresponding incompatible or compatible interaction. (M1) is
 CC useful for conferring resistance to resistance or tolerance to a plant to
 CC bacterial, fungal or viral infection. The present sequence was used to
 CC illustrate the invention.

XX Sequence 2000 3P; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

XX Query Match 6.3%; Score 37.8; DB 7; Length 2000;

XX Best Local Similarity 9.4%; Pred. No. 0.1; Mismatches 228; Indels 2; Gaps 1;

XX Matches 48; Conservative 231; Mismatches 228; Indels 2; Gaps 1;

XX 16 GACAGCAACCTTACGACCTGCTGATACCGACGCTCTAGCTCTAAAGTGTCTGACAG 75

XX 1009 SMSWTYAMWKYTKYMTAYSTWKMWYAMRSRKTWCTGKEMATYGTW 950

XX 76 AAGCTTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 135

XX 949 AAGRMWRMRAKMYCCMWMKMTSCMWMKMYRTMGMARVYVMMRRRTY 890

XX 136 ATGAGAGTGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 195

XX 889 KMSWRMYMTWTKWATWMTWCMAMWATWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTW 830

XX 196 GTTCAGCGGCGATTCAGACGCTATATTTTAAATGTTGGAGAGAGGAATGGTGA 255

XX 829 MKKMTWTTACAMRATSRMRAAGMAMKMYKMKRAVWMTWMTWMTWMTWMTWMTWMTWMT 770

XX 256 GAAACAGAGGCTGACCTTAAAGTACTGACGCGGTGCGCTCTCTCACTAGCTTACT 315

XX 769 KXATATYKMMAMTMMWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMT 710

XX 316 AATTAACCTTGAAGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 373

XX 709 KAKRSYRFRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYR 650

XX 374 CGATTAAGCAAAATTAACCAATCAACGCGCGCGGGAATGGGATTCAGCTGAG 433

XX 649 MMYSKTSCAKCKCKRMYTSSMTGMYTSSMTGMYTSSMTGMYTSSMTGMYTSSMTGMYTSS 590

XX 434 ATTCGACTTACTGCAAGATTTCTGGTACTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCT 493

XX 589 STRRSZMGMSGMSMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYR 530

XX 494 GTAGAAAGATGTCGCGAGATGATGCT 522

XX 529 YTGRTYKATYTSKRYMYKRYKMYWY 501

XX Db

XX RESULT 6

XX ABN67902 standard; DNA; 1332 BP.

XX ABN67902;

XX 01-UTL-2002 (first entry)

XX Streptococcus polynucleotide SEQ ID NO 3717.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

XX KM group A streptococcus; Streptococcus pyogenes; antibacterial; gene;

XX antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.

XX Streptococcus agalactiae.

XX MO200234771-A2.

XX 02-MAY-2002.

XX 29-OCT-2001; 2001WO-GB004789.

XX 27-OCT-2000; 2000GB-00026333.

XX 24-NOV-2000; 2000GB-00028727.

XX 07-MAR-2001; 2001GB-00005640.

XX (CHIR-) CHIRON SPA.

XX (GENO-) INST GENOMIC RES.

XX Telford J, Massignani V, Margarit Y Rostl, Grandi G, Fraser C;

XX Tectelin H;

XX WPI; 2002-352536/38.

XX P-PSDB; ABP27271.

XX New Streptococcus protein for the treatment or prevention of infection or

XX disease caused by Streptococcus bacteria, such as meningitis, and for

XX detecting a compound that binds to the protein.

XX Claim 7; Page 3526; 4525p; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B

XX streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS

XX (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in

XX the specification. The proteins have antibacterial and anti-inflammatory

XX activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and

XX antibodies that bind (I) are used in the manufacture of medicaments for

XX the treatment or prevention of infection or disease caused by

XX Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.

XX Nucleic acids encoding (I) are used to detect Streptococcus in a

XX biological sample. (I) is used to determine whether a compound binds to

XX (I). A composition comprising (I) or a nucleic acid encoding (I), may be

XX used as a vaccine or diagnostic composition. The disease caused by

XX Streptococcus that is prevented or treated may be meningitis. Nucleic

XX acid encoding (I) may be used to recombinantly produce (I) and may be

XX used in gene therapy. Antibodies to (I) are used for affinity

XX chromatography, immunoassays, and distinguishing/identifying

XX Streptococcus proteins

XX Sequence 1332 BP; 423 A; 205 C; 278 G; 426 T; 0 U; 0 Other;

XX Query Match 6.2%; Score 36.8; DB 6; Length 1332;

XX Best Local Similarity 51.9%; Pred. No. 0.18; Mismatches 97; Indels 3; Gaps 1;

XX Matches 108; Conservative 0; Mismatches 97; Indels 3; Gaps 1;

XX 55 AGCTTAATGCTGTGCAAGAGACCTTTATCTGCTAAAGTGAATAAATGTTGGTGCAG 114

XX 517 ACTTTGATGATGCTGCAAGAAATGACACGATGATGATGATGATGATGATGATGATGATGAT 576

XX 115 AAAGGGGCTCTTACAGAGGATATGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 174

XX 577 AGTGAAGAACTTTGCAAGCTGAAAGAAATGAGATGCTGCAAGGATTTCTGCTAGAT 636

XX 175 GCAAGAACACAGCTTC--TAAGCTTACGCGGACATTCAGAGCTTATATTTTCAAT 231

XX 637 GAAATGATGCGCGCTGAGTATATGTTATGTTATGTTATGTTATGTTATGTTATGTTATGTT 696

XX 232 AGTTTCGAGAGAGAGATGTTAGTAA 259

XX 697 AATGATGACATCAATTAATTTGAAGGAA 724

XX Db

XX RESULT 7

XX AN71527_13/c

XX Continuation (14 of 22) of AN71527 from base 130001 (Streptococcus polynucleotide SEQ 1

XX WP Sequence split into 22 fragments LOCUS AN71527 Accession An71527

XX Fragment Name Begin End

XX AN71527_00 1 110000

XX AN71527_01 100001 210000

XX AN71527_02 200001 310000

XX AN71527_03 300001 410000

XX AN71527_04 400001 510000

WP ABN71527_05 500001 610000
 WP ABN71527_06 600001 710000
 WP ABN71527_07 700001 810000
 WP ABN71527_08 800001 910000
 WP ABN71527_09 900001 1010000
 WP ABN71527_10 1000001 1110000
 WP ABN71527_11 1100001 1210000
 WP ABN71527_12 1200001 1310000
 WP ABN71527_13 1300001 1410000
 WP ABN71527_14 1400001 1510000
 WP ABN71527_15 1500001 1610000
 WP ABN71527_16 1600001 1710000
 WP ABN71527_17 1700001 1810000
 WP ABN71527_18 1800001 1910000
 WP ABN71527_19 1900001 2010000
 WP ABN71527_20 2000001 2110000
 WP ABN71527_21 2100001 2155561

Query Match 6.2%; Score 36.8; DB 6; Length 110000;
 Best Local Similarity 51.9%; Pred. No. 1.7; Indels 3; Gaps 1;
 Matches 108; Conservative 0; Mismatches 97

QY 55 AGTCTAATGATGTCGACGAGAGCTTTATCTGCTGAAGTTGAAAAAGTTGGTGCAG 114
 DB 76027 ACTTTGATGATGCTGACGAGAGATGACACGCTGATGAAATTTGATTTGACGAGAT 75968
 QY 115 AAAAGGGGCTCTTAACGAGGGTATGAGAGTGTGCTGCTACTCTTTACGCACTGCG 174
 DB 75967 AGTGAAGAACTTTGGAAGCTGGAAGAAATGAGATGCTGCAAGGAGATTTCTGCTAGAT 75908
 QY 175 GCAGAGAACACAGTCTCC---TAAGTTCAGCGGCGAGATTCAGAGCTTATTTCAAT 231
 DB 75907 GAAATGATGCGCGGCTGAAGTTATGTTCCGCGCACTGATCTTCTGATGATCAATCAAC 75848
 QY 232 AGTTTCGAGAGAGGAATGTGTAGTAA 259
 DB 75847 AATGATGCAATCAATATTTGAAGAA 75820

RESULT 8
 ACA38387/c
 ID ACA38387 standard; DNA; 1605 BP.
 AC ACA38387;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Prokaryotic essential gene #20044.
 XX
 KM Antisense; ds; prokaryotic essential gene; cell proliferation;
 XX drug design; gene.
 XX
 OS Mycobacterium bovis.
 XX
 PN MO200277183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948933.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Marg L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW,
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth KA, Xu HH;
 XX
 DR WPI; 2003-029926/02.
 DR P-PSDB; ABU34517.

XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.

Claim 14, SEQ ID NO 26257, 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of the 623 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-regulated gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 1605 BP; 334 A; 490 C; 516 G; 265 T; 0 U; 0 Other;

Query Match 6.1%; Score 36.6; DB 7; Length 1605;
 Best Local Similarity 57.4%; Pred. No. 0.23; Indels 0; Gaps 0;
 Matches 66; Conservative 0; Mismatches 49

QY 184 ACGTCTCCTAAGTTCAGCGCGGAGATTGACGCTTATTTCAATTAATTTGAGAG 243
 DB 200 ACGGTATTTGCGCGGAGATGACGCTGAGCTCGCATTTGTTGGTTCTCGGACAG 141
 QY 244 AGAATGTGTAGTAAACAGAGGTTGACCTTAAAGAGTCTGACGGGTGTGCC 298
 DB 140 GGTGGTGTGTAGGAGACAGATTGTGTTACTGACGAGCTACTGACGCGGTACACGC 86

RESULT 9
 AA199682_18/c
 Continuation (19 of 45) of AA199682 from base 1800001 (Mycobacterium tuberculosis strain
 WP Sequence split into 45 fragments LOCUS AA199682 Accession AA199682
 WP Fragment Name Begin End
 WP AA199682_00 1 110000
 WP AA199682_01 100001 210000
 WP AA199682_02 200001 310000
 WP AA199682_03 300001 410000
 WP AA199682_04 400001 510000
 WP AA199682_05 500001 610000
 WP AA199682_06 600001 710000
 WP AA199682_07 700001 810000
 WP AA199682_08 800001 910000
 WP AA199682_09 900001 1010000
 WP AA199682_10 1000001 1110000
 WP AA199682_11 1100001 1210000
 WP AA199682_12 1200001 1310000
 WP AA199682_13 1300001 1410000

```

WP AA199682_14 1400001 1510000
WP AA199682_15 1500001 1610000
WP AA199682_16 1600001 1710000
WP AA199682_17 1700001 1810000
WP AA199682_18 1800001 1910000
WP AA199682_19 1900001 2010000
WP AA199682_20 2000001 2110000
WP AA199682_21 2100001 2210000
WP AA199682_22 2200001 2310000
WP AA199682_23 2300001 2410000
WP AA199682_24 2400001 2510000
WP AA199682_25 2500001 2610000
WP AA199682_26 2600001 2710000
WP AA199682_27 2700001 2810000
WP AA199682_28 2800001 2910000
WP AA199682_29 2900001 3010000
WP AA199682_30 3000001 3110000
WP AA199682_31 3100001 3210000
WP AA199682_32 3200001 3310000
WP AA199682_33 3300001 3410000
WP AA199682_34 3400001 3510000
WP AA199682_35 3500001 3610000
WP AA199682_36 3600001 3710000
WP AA199682_37 3700001 3810000
WP AA199682_38 3800001 3910000
WP AA199682_39 3900001 4010000
WP AA199682_40 4000001 4110000
WP AA199682_41 4100001 4210000
WP AA199682_42 4200001 4310000
WP AA199682_43 4300001 4410000
WP AA199682_44 4400001 441529

```

Query Match Best Local Similarity 57.4%; Score 36.6; DB 4; Length 110000;
Matches 66; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

```

OY 184 ACGTCTCTTAAGTTCACGCGCGAGATTCAGACGTTATATTTCAATAAGTTTCGAGAG 243
DB 33577 ACGGCTACTTCCGCGCGAGGTACGCTGGGACTCGCATATTTGGTTCTCCGACAG 33518
OY 244 AGGAATGTGTAGTACAGAGGGTGACCTTAAGAGGTACTCGACGGGTGTGGC 298
DB 33517 GGTGGGTGTGTAGGAGACAGATTGTGTAAGTCTGACGAGGCTACTCGACGGGTACAGC 33463

```

RESULT 10
AA199683_18/c
Continuation (19 of 44) of AA199683 from base 1800001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments Locus AA199683 Accession AA199683
Fragment Name Begin End
WP AA199683_00 1 110000
WP AA199683_01 100001 210000
WP AA199683_02 200001 310000
WP AA199683_03 300001 410000
WP AA199683_04 400001 510000
WP AA199683_05 500001 610000
WP AA199683_06 600001 710000
WP AA199683_07 700001 810000
WP AA199683_08 800001 910000
WP AA199683_09 900001 1010000
WP AA199683_10 1000001 1110000
WP AA199683_11 1100001 1210000
WP AA199683_12 1200001 1310000
WP AA199683_13 1300001 1410000
WP AA199683_14 1400001 1510000
WP AA199683_15 1500001 1610000
WP AA199683_16 1600001 1710000
WP AA199683_17 1700001 1810000
WP AA199683_18 1800001 1910000
WP AA199683_19 1900001 2010000
WP AA199683_20 2000001 2110000
WP AA199683_21 2100001 2210000
WP AA199683_22 2200001 2310000

```

WP AA199683_23 2300001 2410000
WP AA199683_24 2400001 2510000
WP AA199683_25 2500001 2610000
WP AA199683_26 2600001 2710000
WP AA199683_27 2700001 2810000
WP AA199683_28 2800001 2910000
WP AA199683_29 2900001 3010000
WP AA199683_30 3000001 3110000
WP AA199683_31 3100001 3210000
WP AA199683_32 3200001 3310000
WP AA199683_33 3300001 3410000
WP AA199683_34 3400001 3510000
WP AA199683_35 3500001 3610000
WP AA199683_36 3600001 3710000
WP AA199683_37 3700001 3810000
WP AA199683_38 3800001 3910000
WP AA199683_39 3900001 4010000
WP AA199683_40 4000001 4110000
WP AA199683_41 4100001 4210000
WP AA199683_42 4200001 4310000
WP AA199683_43 4300001 4403765

```

Query Match Best Local Similarity 57.4%; Score 36.6; DB 4; Length 110000;
Matches 66; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

```

OY 184 ACGTCTCTTAAGTTCACGCGCGAGATTCAGACGTTATATTTCAATAAGTTTCGAGAG 243
DB 24455 ACGGCTACTTCCGCGCGAGGTACGCTGGGACTCGCATATTTGGTTCTCCGACAG 24396
OY 244 AGGAATGTGTAGTACAGAGGGTGACCTTAAGAGGTACTCGACGGGTGTGGC 298
DB 24395 GGTGGGTGTGTAGGAGACAGATTGTGTAAGTCTGACGAGGCTACTCGACGGGTACAGC 24341

```

RESULT 11
ABX74308
ID ABX74308 standard; DNA; 593 BP.
XX
AC ABX74308;
XX
DT 18-MAR-2003 (first entry)
XX
DE DNA sequence #80 encoding M. tuberculosis antigenic polypeptide.
XX
KM Mycobacterium tuberculosis antigenic polypeptide; immune response;
XX tuberculosis infection; gene therapy; gene; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN US6465633-B1.
XX
PD 15-OCT-2002.
XX
PF 23-DEC-1999; 99US-00470191.
XX
PR 24-DEC-1998; 98US-0113952P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skelky Y;
XX
DR WPI, 2003-147072/14.
XX
XX Novel isolated mycobacterial polynucleotide, useful for treating,
PT preventing or diagnosing Mycobacterium tuberculosis infection, for
PT producing Mycobacterium tuberculosis secretory polypeptides and DNA
PT vaccines.
PS Example; Col 81-82; 48pp; English.
XX
CC The present invention relates to the isolation of polynucleotide
sequences encoding Mycobacterium tuberculosis antigenic polypeptides. The

CC polynucleotide sequences of the invention are useful for treating,
CC preventing, and diagnosing M. tuberculosis infection, for producing M.
CC tuberculosis secretory polypeptides, for producing DNA vaccines, for
CC diagnostic purposes, as molecular probes or primers to detect the
CC presence of bacteria in a biological sample, for inducing and/or
CC enhancing immune responses to M. tuberculosis, and in gene therapy.
CC ABX74229-ABX74319 represent DNA sequences encoding M. tuberculosis
CC antigenic polypeptides
XX
SQ Sequence 593 BP; 95 A; 181 C; 186 G; 120 T; 0 U; 11 Other;
Query Match 6.0%; Score 35.6; DB 7; Length 593;
Best Local Similarity 56.5%; Pred. No. 0.31;
Matches 65; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 184 ACGGCTCTTAAGTTCAGCGCGGCGGAGTTCAGCGTTATTTCAATAGTTTCGAGAG 243
DB 101 ACGGCTCTTAAGTTCAGCGCGGCGGAGTTCAGCGTTATTTGTTGCTTCTCCGACAG 160
QY 244 AGGAATGTGTAGTACAGAGGAGTACCTTACAGAGTACGAGGAGTGTGCGC 298
DB 161 GTTGTGCTGTACGAGGAGTGTACCTTACAGAGGAGTACGAGGAGTACAGGC 215
RESULT 12
AAL04577
ID AAL04577 standard; DNA; 6756 BP.
XX
AC AAL04577;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 7265.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
XX
OS Homo sapiens.
XX
XX WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001339.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 11-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227003P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229341P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 14-SEP-2000; 2000US-0233066P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235844P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235835P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 29-SEP-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239337P.
PR 13-OCT-2000; 2000US-0239337P.
PR 13-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0244674P.
PR 08-NOV-2000; 2000US-0244745P.
PR 08-NOV-2000; 2000US-0244746P.
PR 08-NOV-2000; 2000US-0244776P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.

PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
(HUMA-) HUMAN GENOME SCI INC.
XX PA
XX PA
XX PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465570/50.
XX
XX PT Isolated nucleic acid molecule encoding a reproductive system antigen is
XX used in preventing, treating or ameliorating a medical condition.
XX
XX PS
XX PS Disclosure; SEQ ID NO 7265; 1297bp + Sequence Listing; English.
XX
XX CC The present invention provides the protein and coding sequences of a
XX number of human reproductive system related antigens. These can be used
XX in the prevention and treatment of reproductive system disorders,
XX including cancer. The present sequence is a genomic sequence encoding a
XX protein of the invention
XX
SQ Sequence 6756 BP; 1603 A; 1601 C; 1877 G; 1675 T; 0 U; 0 Other;
Query Match 5.7%; Score 33.8; DB 4; Length 6756;
Best Local Similarity 56.9%; Pred. No. 4.3;
Matches 62; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
QY 218 TATATTTTCAATAGCTTTCGAGAGAGAGATGTGTAGTAACAGAGGTCACCTTAAGA 277
DB 1737 TATATTTTTCATATTTTATTTAGCAGAAAGAGATGTAGTAGAGAGCGGTGATGATAGG 1796
QY 278 AGGTACTCGACGGGTGGCGCCTTCAGTACAGTTCCTACTATAACTAG 326
DB 1797 AGAGAGTCTGCAGAGACATGTGACGAAAGAAATCTATGTCTATATTAG 1845
RESULT 13
ABL97500
ID ABL97500 standard; DNA; 6756 BP.
XX
AC ABL97500;

XX DT 21-JUN-2002 (first entry)
XX DE Human testicular antigen encoding DNA fragment SEQ ID NO: 2152.
XX
XX KW Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KW reproductive system disorder; urinary system disorder; gene therapy;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disease; infection; cytostatic; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200155317-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001329.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-019076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232081P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.

```

PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235835P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241755P.
PR 20-OCT-2000; 2000US-0241756P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246529P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249267P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.

```

```

PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI, 2001-483232/52.
XX
XX Nucleic acids encoding 973 human testicular antigen polypeptides, useful
XX for preventing, diagnosing and/or treating testicular cancer.
XX
XX PS Disclosure; SEQ ID NO 2152; 766pp; English.
XX
XX CC The present invention provides the protein and coding sequences of 973
XX human testicular antigens, and fragments of their genomic sequences. The
XX sequences can be used in the treatment of cardiovascular, urinary system,
XX reproductive system, immune, respiratory, neurological and
XX gastrointestinal disorders, infections, and particularly cancer,
XX especially testicular cancers. The present sequence is a DNA encoding a
XX protein fragment of the invention
XX
XX SQ Sequence 6756 BP; 1603 A; 1601 C; 1877 G; 1675 T; 0 U; 0 Other;
XX
XX Query Match 5.7%; Score 33.8; DB 4; Length 6756;
XX Best Local Similarity 56.9%; Pred. No. 4.3;
XX Matches 62; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
XX
XX QY 218 TTATATTTTCAATAGTTCGAGAGAGAGATGCTGTACACAGAGGCTGACTTAGA 277
XX |||||
XX DB 1737 TTATGATTTTCATTTTATTAGAGAAAGAAATGATGAGAGCGGATGATGATTAAGG 1796
XX |||||
XX QY 278 AGGTACTGCACGGGTGTCGCCCTCTCACTAGATTCTACTATATAAATTAG 326
XX |||||
XX DB 1797 AAGAGCTCGCAGAGACATGTGACCAAAAGATCTATGCTATATTAG 1845
XX |||||
XX
XX RESULT 14
XX AA182806/c
XX ID AA182806 standard; cDNA; 360 BP.
XX
XX AC AA182806;
XX
XX DT 06-NOV-2001 (first entry)
XX
XX DE Human polynucleotide SEQ ID NO 2866.
XX
XX XX Human polynucleotide SEQ ID NO 2866.
XX
XX KM Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorders; arthritis; inflammation; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200164835-A2.
XX
XX PD 07-SEP-2001.
XX
XX PF 26-FEB-2001; 2001WO-US0049927.
XX
XX XX 28-FEB-2000; 2000US-00515126.
XX
XX PR 18-MAY-2000; 2000US-00577409.
XX
XX XX (HYSE-) HYSEQ INC.
XX
XX PA Tang YT, Liu C, Dymnac RT;
XX
XX PI
XX

```

DR WPI: 2001-514838/56.
 DR P-PSDB: AAO02875.
 XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing
 PT and treating e.g. leukemia, inflammation and immune disorders.
 XX
 PS Claim 1, SEQ ID NO 2866; 1339pp + Sequence listing; English.
 CC The invention relates to human polynucleotides (AA179941-AA193841) and
 CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activity/inhibit activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 360 BP; 89 A; 70 C; 106 G; 95 T; 0 U; 0 Other;
 Query Match 5.6%; Score 33.2; DB 4; Length 360;
 Best Local Similarity 50.6%; Pred. No. 1.6; 78; Indels 0; Gaps 0;
 Matches 80; Conservative 0; Mismatches 0;
 QY 239 GAGGAGGAGATGTGTAGTACAGAGGGTACCTTAGAAGTACTGACGGGTGTGCGC 298
 Db 203 GACACGAAATATATCTAGAAAGGCAAGATGCTCCCAAGTAGAGAGCCCTGGTTGGAT 144
 QY 299 CTCTCAGTACTACTAATAAATTAGAAGCTTCGCTACTTTCACCTGAGGCTTAA 358
 Db 143 CCCAAGCAAGTACTTTCACCTTCGACCATTTCCCTGTCTTTAACTGAGGCTTACA 84
 QY 359 TTGACTTTGTATCGCGTATAGACAAATTACCCCA 396
 Db 83 TTTAAATGCTATCTTATAGATCAATTACCCATA 46
 RESULT 15
 AAD14902/c
 ID AAD14902 standard; cDNA; 972 BP.
 AC AAD14902;
 XX 01-NOV-2001 (first entry)
 DE Human transporter-related protein #15 cDNA.
 XX Human; transporter-related protein; metabolite transporter;
 KM organic cation transporter; multi-drug resistance; MDR; gene therapy;
 KM cellular transporter; sodium-glucose cotransporter; diagnosis; screening;
 KM symptomatic; phenotypic manifestation; biological condition; ss.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..972
 FT /tag= a
 FT /product= "Human transporter-related protein"
 FT
 XX MO200157214-A2.
 XX
 XX 09-AUG-2001.
 PD
 XX 02-FEB-2001; 2001WO-US003646.
 PF
 XX 03-FEB-2000; 2000US-0179973P.
 PR 14-FEB-2000; 2000US-0182422P.
 XX
 PA (LEXI-) LEXICON GENETICS INC.

XX Turner CA, Mathur B, Wang X, Abuin A, Friedrich G, Zambrowicz B,
 PI Sands AT;
 XX WPI: 2001-514599/56.
 DR P-PSDB: AAE08068.
 XX
 PS Novel polynucleotides encoding novel human proteins with structural
 PT similarity to cellular transporters for the diagnosis of disease and use
 PT in gene therapy.
 XX
 PS Disclosure; Page 53; 91pp; English.
 CC The present sequence is a human transporter-related protein cDNA. The
 CC human transporter-related protein share structural similarity with
 CC mammalian metabolite or organic cation transporters, multi-drug
 CC resistance (MDR) proteins, mammalian sodium-glucose cotransporters and
 CC other cellular transporters. The transporter-related protein DNA may be
 CC used for the detection of mutant sequences or inappropriately expressed
 CC sequences for the diagnosis of disease. They may also be used to screen
 CC for drugs effective in the treatment of the symptomatic or phenotypic
 CC manifestations of perturbing the normal functions of the sequences of the
 CC invention in the body. They may also be used in gene therapy for treating
 CC biological conditions
 XX
 SQ Sequence 972 BP; 192 A; 257 C; 242 G; 280 T; 0 U; 1 Other;
 Query Match 5.5%; Score 32.6; DB 4; Length 972;
 Best Local Similarity 53.5%; Pred. No. 4.2; 59; Indels 0; Gaps 0;
 Matches 68; Conservative 0; Mismatches 0;
 QY 453 TTTCTGGGTACTGCGCCGAAGCTCTCTGAATTAACAGCAAGTAGAAGATGTTCGCGAG 512
 Db 954 TGTCTTATAGTCTCTCATTAATCTACACAGCAAGTAAGTGAAGTGTAAACCA 895
 QY 513 TATGACCTCTAAAAGTGAAGGTGAGTGTAAATACCACTGAGCAATCTGCGTCA 572
 Db 894 CAGCAAGACTCTCTGGGATATTAATGAGGCCACCATTTGACCAATGGACACAGGAGCC 835
 QY 573 GCTAGGT 579
 Db 834 GCTGTT 828

Search completed: April 6, 2004, 23:56:11
 Job time : 307 secs

more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0A1040CF04NP1&cluster=8230.r. Contact :
Peng Liang Email : filiang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0A1040CF04NP1.

FEATURES
source
1. .1103
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1040Y107"
/feature="CS0D1040Y107" COT 25-NORMALIZED

FEATURES
source
Location/Qualifiers
1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1022YF05"
/feature="CS0D1022YF05" COT 25-NORMALIZED
/note="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

Fri Apr 9 15:30:05 2004

US-09-613-486-14.rst

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 23:23:08 ; Search time 2367 Seconds
(without alignments)
7531.779 Million cell updates/sec

Title: US-09-613-486-14

Perfect score: 597
Sequence: 1 atggagctgctgcgcacag.....gtagaagggaagttatgttaa 597

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931080276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estm:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_hnc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_est3:*
- 12: gb_est4:*
- 13: gb_est5:*
- 14: gb_estfun:*
- 15: em_estom:*
- 16: em_estom:*
- 17: em_gss_hum:*
- 18: em_gss_inv:*
- 19: em_gss_pln:*
- 20: em_gss_vtc:*
- 21: em_gss_fun:*
- 22: em_gss_mam:*
- 23: em_gss_mus:*
- 24: em_gss_pro:*
- 25: em_gss_rnd:*
- 26: em_gss_dhg:*
- 27: em_gss_vr1:*
- 28: gb_gss1:*
- 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

5	36.4	6.1	1201	13	EX416567	EX416567
6	35.8	6.0	507	9	AA530213	AA530213
7	35.8	6.0	545	14	CB166256	CB166256
8	35.6	6.0	1201	13	EX395102	EX395102
9	35.4	5.9	1201	9	AL556952	AL556952
10	35.2	5.9	511	28	AO562700	AO562700
11	35.2	5.9	1106	29	CNS001RE	AL078654
12	35.2	5.9	1201	13	EX340305	EX340305
13	35	5.9	460	9	AA712605	AA712605
14	35	5.9	814	28	BH177584	BH177584
15	35	5.9	814	29	CNS07K3H	AL614543
16	35	5.9	830	10	BP215190	BP215190
17	35	5.9	838	29	CNS0256D	AL181678
18	34.8	5.8	581	29	PR0014687	AL005896
19	34.8	5.8	708	28	BH933072	BH933072
20	34.8	5.8	1201	13	EX394291	EX394291
21	34.6	5.8	850	29	CNS036YG	AL230641
22	34.4	5.8	361	12	BI050737	BI050737
23	34.4	5.8	508	10	BE141051	BE141051
24	34.4	5.8	938	13	BO882089	BO882089
25	34.4	5.8	950	13	BP720645	BP720645
26	34.4	5.8	1201	9	AL534306	AL534306
27	34	5.7	650	12	BP776877	BP776877
28	33.8	5.7	552	29	CB369513	CB369513
29	33.8	5.7	657	13	BQ234878	BQ234878
30	33.8	5.7	690	12	BG535075	BG535075
31	33.6	5.6	444	10	BF595177	BF595177
32	33.6	5.6	426	9	AI900854	AI900854
33	33.6	5.6	506	12	BP790558	BP790558
34	33.6	5.6	519	14	CA785448	CA785448
35	33.6	5.6	588	13	BQ094372	BQ094372
36	33.6	5.6	848	29	CNS03USZ	AL261548
37	33.4	5.6	428	14	CB794203	CB794203
38	33.4	5.6	690	13	BM300017	BM300017
39	33.4	5.6	830	28	AO367141	AO367141
40	33.4	5.6	818	29	CNS0202P	AL176254
41	33.4	5.6	952	12	BI079611	BI079611
42	33.4	5.6	1201	13	EX461787	EX461787
43	33.2	5.6	703	28	AQ403529	AQ403529
44	33.2	5.6	760	28	AZ732303	AZ732303
45	33.2	5.6	831	10	BF577284	BF577284

ALIGNMENTS

RESULT 1
LOCUS BE915548/c
DEFINITION 601668490F1 NCI_GAP_Mam1 Mus musculus CDNA clone IMAGE:3968231 5',
mRNA sequence.
ACCESSION BE915548
VERSION BE915548.1 GI:10415505
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 693)
REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

Query Match	6.8%;	Score 40.4;	DB 10;	Length 693;
Best Local Similarity	52.4%;	Pred. NO. 0.27;		

us-09-613-486-14.rst

Page 4

Email: cgapops-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hnsc.bcm.tmc.edu/cdna/>
 Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Huiyk, S.W., Louisgeed, H.,
 Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Narasvati,
 A.N., Gibbs, R.A.

FEATURES

Location/Qualifiers

```

source
1..1201
/organism="Homo sapiens"
/mol_type="cDNA"
/db_xref="taxon:9606"
/clone="CSOD008YH20"
/tissue_type="NEUROBLASTOMA"
/clone_id="Homo sapiens NEUROBLASTOMA"
/notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."

```

ORIGIN

Query Match	6.1%;	Score 36.4;	DB 13;	Length 1201;
Best Local Similarity	15.3%;	Pred. No. 7;		
Matches 41;	Conservative 88;	Mismatches 138;	Indels 1;	Gaps 1;

Oy	98	AAAAATGTGTGCAGAAAGGGCTCTTAACGAGGATATAGAAATGTGTGGCTTAC	157
Db	475	ANNNAANNKGGGGTGCAAGGNNNNNNNNAANGGGGNNGGGNNKNGNNGNN	534
Oy	158	TCCTTTACGACTCGCGGAGAAACACAGTCTCTTAAGTTGAGCGCGCAGATTTCAGCG	217
Db	535	NTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	594
Oy	218	TTATATTTTCAAATAGTTTCGAGAGAGAAATGTGTAGTAAACAGAGGTACCTTAA	277
Db	595	CTKKKTKNNCMABETKTKKKKTKTKKKKTKAAMKMMCMCMMSMGVKKGGSVSCAAMV	654
Oy	278	AGGTACTCGAGGGGTGTGGCCTCTCACTAGGTTCACTAATTAATTGAAGAGTTGGCT	337
Db	655	CGGVSDIMCMCKTKGKGNMMCKMKCMKKGGKHGTH--MMCTKHHMYTKTBACXB	713
Oy	338	GTACTTCACTGAGGCTTACGTTGACTT	365
Db	714	MTKYKCTWCTNKTKACTMMEKKYKVK	741

RESULT	6
AA530213	
LOCUS	
DEFINITION	507 bp mRNA linear EST 22-JUL-1997
IMAGES:	vj4ch05.r1 Striatagene mouse sk:n (#937313) Mus musculus CDNA clone
IMAGE:	932121 5' mRNA sequence.
AA530213	
VERSION	AA530213.1 GI:2272919

TITLE	Waterston, R.
JOURNAL	The Washu-HHI Mouse EST Project
COMMENT	Unpublished (1996)
CONTACT	Contact: Marra M/Mouse EST Project
REMARKS	Waterston, R. (1996) Mouse EST Project

WashU-RHMI Mouse B61 Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LINT; contact the
IMaG Consortium (info@image.lnlnl.gov) for further information.
MGI:533741
Seq primer: -28m3 rev1.5T from Amersham
High quality sequence, stop: 162.

FEATURES

SOURCE

```

/organism="mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:93121"
/sex="females"
/tissue_type="whole skin"
/dev_stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/clone_id="Stratagene mouse skin (#937313)"
/note="Organ: skin; Vector: pBluescript SK-; Site: 1:
ECOR1; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Whole skin from 11 week old C57BL/6 female mice
Average insert size: 1.0 kb; UniZAP XR Vector: ~5'
adaptor sequence: 5' GAATTCGACGACG 3' ~3' adaptor
sequence: 5' CTCGAGTTTATTTTATTTT 3'"

```

ORIGIN

Query Match	6.0%;	Score 35.8;	DB 9;	Length 507;
Best Local Similarity	51.6%;	Fred, No. 6.7;		
Matches 82;	Conservative 0;	Mismatches 77;	Indels 0;	Gaps 0

QY 30 CAACCTGTGATMACCGACGCGCTTACGTCTAAATGTGTGACACAAAGCTTTATCTCG 89
Db 300 CATCGCTGTGCTGTCTACCTTGAAGTCGAAATGATTCACAGTTGGCGCTTCAATATA 359
QY 90 TGAAGTTAAAAAAGTTGTGTGCAAGAAAGGGGCTCTTAACGAGGTAATGAATGGTGT 149
Db 360 TCAGGGGAACGCTGTTCCGCGAGTAAAGGCACTGTGATGTGAGGAAACAAACAGTGT 419
QY 150 CGGTCTACTCCTTTACGCACTCGCGGCAAGAACACGCTC 188
Db 420 CAGCGGTACCCATCAGCCATCAATATGACAGGTCAACACGTC 458

RESULT 7					
LOCUS	CB166256				
DEFINITION	CB166256	545 bp	mRNA	linear	EST 30-JAN-200
ACCESSION	XAB602600277.R1	CSE0FXL03	Kidney Bos taurus	CDNA	mRNA sequence.
VERSION	CB166256.1	GI:28152381			
KEYWORDS	EST.				
SOURCE	Bos taurus (cow)				
ORGANISM	Bos taurus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;				

Tel: 9798452616
Fax: 9798456970
Email: david.adelson@amu.edu.
Location/Qualifiers

FEATURES

SOURCE

1..545
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/issue_type="kidney"
/clone_lib="CEGFXL03 kidney"
/note="Organ: kidney; Vector: pBluescript SK+; Site_1: NotI; Site_2: EcoRI; sequence 5' of the insert (5'-NNN...NNNinsert)
GCGAATGAGCTCCACCGCGGTGGCGCGCGCGCTCGAG. Sequence 3' of the insert (AAGATTGATATCAAGCTTATGATACCTGACCTGAG. non-normalized library, sequenced 3' with M13R primer."

ORIGIN

Query Match 6.0%; Score 35.8; DB 14; Length 545;
Best Local Similarity 52.3%; Pred. No. 7;
Matches 79; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 430 GAAGATTGCTACTAGCTGAGATTTCTGGTACTTCCCGAAGCTCTCTGAATTACAG 489
DB 94 GATGCTGATAGAGTGAATGATACATTTATGATGATATTTCCCTCAGTCCGCTACAAACCTT 153
QY 490 CAAGTGAAGAAGATGTCGCGAGTATGATGCTTAAACCTGAGGTGAGTGTAAAT 549
DB 154 CAAGAAGGAGATCAATGCTCTGTTTATAGAGAGAGAACTGAACTTACGGAGAGTCA 213
QY 550 ACACGAGGAGCATCTGCGTACAGTACAGTA 580
DB 214 ACAGATTGAGGATCAAGTACAGCTTACGAA 244

RESULT 8
BX395102 1201 bp mRNA linear EST 13-MAY-2003
LOCUS BX395102 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
DEFINITION cDNA clone CS0DC02YK20 5-PRIME, mRNA sequence.
ACCESSION BX395102.1 GI:30616562
VERSION BX395102.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS L.W.B., Gruber,C., Jessee,J., and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
BP 191 91006 EVRY cedex - France
Email: secrete@genoscope.cns.fr; Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 2431.f For more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DC029BP100P1&cluster=2431.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradey Avenue Genoscope sequence ID : CS0DC029BP100P1.
Location/Qualifiers

FEATURES

SOURCE

1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DC029YK20"
/issue_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 6.0%; Score 35.6; DB 13; Length 1201;
Best Local Similarity 45.2%; Pred. No. 13;
Matches 57; Conservative 17; Mismatches 52; Indels 0; Gaps 0;

QY 211 TCAGACGTTATATTTCAATAGTTTCGAGAGAGAAATGTGTAGTAACAGAGGTGAC 270
DB 1071 TCAGAGTTATATTTGTAATTAATAATATGAKATGAGCGYKTTATAGGKTGTT 1130
QY 271 CTTAAGAAGTACTGAGCGGTGCGGCTCTCACTAGTTCATTAATTAAGTACAG 330
DB 1131 CTTGAAACCHYTWCTYTGTCWGGGCTTAAVCTCCCAATTAAGACATTCATG 1150
QY 331 TTGCGT 336
DB 1191 TTGKMT 1196

RESULT 9
AL556952 1201 bp mRNA linear EST 31-MAY-2003
LOCUS AL556952 Homo sapiens T CELLS (JURKAT CELL LINE) Homo sapiens cDNA
DEFINITION clone CS0DH005YA07 3-PRIME, mRNA sequence.
ACCESSION AL556952.2 GI:31278752
VERSION AL556952.2
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS L.W.B., Gruber,C., Jessee,J., and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On Feb 15, 2001 this sequence version replaced gi:12300090.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: secrete@genoscope.cns.fr; Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 7763.r For more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DH005AA04NP1&cluster=7763.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradey Avenue Genoscope sequence ID : CS0DH005AA04NP1.
Location/Qualifiers

FEATURES

SOURCE

1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DH005YA07"
/issue_type="T CELLS (JURKAT CELL LINE)"
/cell_line="JURKAT CELL LINE"
/clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE)"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Query Match 5.9%; Score 35.4; DB 9; Length 1201;
Best Local Similarity 42.3%; Pred. No. 15;
Matches 85; Conservative 18; Mismatches 98; Indels 0; Gaps 0;

QY 70 GACAAGACCTTTATCTGCTGAAGTGAATAAATGTTGGCAGAAAGGGGCTCTTAC 129
DB 1 KATTAAGVCTATGTTGTAACANAGCTGAGCATGATANGCWCACGAGACTTGCT 60

QY 130 GAGGCTATAGAAAGTGGTTCCTGCTACTCTCTTACGCACTCGCGGAGAACCAACGCTC 189
 DB 61 GTGGGTTTGAAGGAGGTGTAGTAACTCTGNNHGBGGCGCTGTGTGATVTKACAGCTTC 120
 QY 190 CTTAAGCTTACGCGGAGATTCAGACGTTATATTTTCAATAGATTTCGAGAGAGAAAT 249
 DB 121 CTTCCGDBGTGGGGGTGAGGTAGCTGTGTGTATATKATKGBATCAKKGTTGGCTT 180
 QY 250 GTGCTAGTACAGAGGCTGAC 270
 DB 181 GGBAAGKGGCCGAGGGTGC 201

RESULT 10
 A0562700/c
 LOCUS
 DEFINITION HS_2079_B1_B05_T7C CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2079 Col=9 Row=D, genomic survey sequence.
 ACCESSION A0562700
 VERSION A0562700.1 GI:4922171
 KEYWORDS GSS.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 511)
 Mahairas G.G., Wallace J.C., Smith K., Swartzell S., Holzman T., Keller A., Shaker R., Furlong J., Young J., Zhao S., Adams M.D. and Hood L.
 Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
 1049764
 Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: http://www.htsc.washington.edu
 Plate: 2079 row: D column: 9
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 511.
 Location/Qualifiers
 1..511
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="Plate=2079 Col=9 Row=D"
 /sex="male"
 /clone.lib="CIT Approved Human Genomic Sperm Library D"
 /note="Organ: sperm; Vector: pBelobaC11; BAC clones in E-Coli DH10B"

ORIGIN
 Query Match 5.9%; Score 35.2; DB 28; Length 511;
 Best Local Similarity 48.3%; Pred. No. 10;
 Matches 97; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 91 GAAGTGTAAAAAGTTGGTGTGAGAAAGGCGCTCCTACAGAGGTATAGAAAGTGTTC 150
 DB 221 GCAAGTGTACATAGTGTGATGTAAAGAGCGGAGAACTAATATGTATAGAGAGAG 162
 QY 151 GGTCTACTCTTTAGCACTCGCGGAGAAACAAGTCTCTTAAGTTTCAGCGGAGAT 210
 DB 161 GGTGTCTCCAAACCTATACCCAGCACTCTCCCAACCCCTCAATTGACAGACAGAT 102

QY 211 TCAGACGTATATTTTCAATAGTTTCGAGAGAGAAATGTGTAGTACAGAGGCTGAC 270
 DB 101 GCAACATAGAAATGCAAGAGAGCTGCTGCTCAGACCGTGCATGCCGAGGTGCTGA 42
 QY 271 CTTAAGAGGTACTGACGG 291
 DB 41 AGTATCAGGAGTGTGACGG 21

RESULT 11
 CNS00LRE
 LOCUS
 DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BACR48P05 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 AL078654
 AL078654.1 GI:5101944
 GSS.
 Drosophila melanogaster (fruit fly)
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 1106)
 Genoscope.
 Direct Submission
 Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
 Web : www.genoscope.cns.fr
 Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammossier in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the library P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
 Location/Qualifiers
 1..1106
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /clone="BACR48P05"
 /clone.lib="RPCI-98"
 /note="end : T7"

ORIGIN
 Query Match 5.9%; Score 35.2; DB 29; Length 1106;
 Best Local Similarity 15.1%; Pred. No. 16;
 Matches 47; Conservative 132; Mismatches 133; Indels 0; Gaps 0;

QY 131 AGGATATAGAAAGTGTGTGCTGCTACTCTTACGCACTCGCGGAGAACCAACGCTC 190
 DB 714 AMRGTGMMRATWMTWKBTTKTKNNRNDVNDVDTKVMGGYBKMKMCSYTYC 773
 QY 191 CTTAAGTTCAGCGGCAAGATTTCAGCGTTATATTTTCAATAGTTTCGAGAGAGAAATG 250
 DB 774 CCCARDYKCVSGGTCCSCTYMANCNTTGAAMWTDKXVSWTBTMTVNTGSKTNR 833
 QY 251 TGGTGTACAGAGGATGACCTTAAGAAGTACTGACGCGGNGTGCCTCTACTAGT 310
 DB 834 RDBKVDYDGTBEGGKBVCHSHYWMKRWMTDATTDKWTBKSCCSDSCNCMR 893
 QY 311 TCATTAATTAAGTAAAGTGTGCTGCTTTCATCTGAGGCTTACGCTTATCTTTGTA 370
 DB 894 WWYDMMRWADKDHRRSARWVCSADMKTITTTTXYKYKTTWTBTBMYVVGGGTHTTW 953

ORGANISM	Schistosoma mansoni
REFERENCE	Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigoida; Schistosomatoidae; Schistosomatidae; Schistosoma.
AUTHORS	1 (bases 1 to 814)
TITLE	Le Paslier, M.-C., Pierce, R. J., Merlin, F., Hirai, H., Wu, W., Williams, D. L., Johnston, D., Loverde, P. T. and Le Paslier, D. Construction and characterization of a Schistosoma mansoni bacterial artificial chromosome library
JOURNAL	Genomics 65 (2), 87-94 (2000)
MEDLINE	20247247
PUBMED	10783255
COMMENT	Other_GSSs: 010_G_21-21 Contact: Pierce RJ INSERM U 167 Institut Pasteur de Lille 1 rue du Professeur A. Calmette, 59019-Lille, France Tel: (33) (0)3 20877783 Fax: (33) (0)3 20877888 Email: Raymond.Pierce@pasteur-lille.fr CNS sequencing ID=DC0A010AD11BP1 Plate: 010 row: G column: 21 Seq primer: M13 reverse primer Class: BAC ends High quality sequence stop: 814. Location/Qualifiers 1..814 /organism="Schistosoma mansoni" /mol_type="genomic DNA" /strain="Puerto-Rican" /db_xref="taxon:6183" /clone="010G21" /sex="mixed" /dev_stage=" cercariae" /lab_host="Biomphalaria glabrata" /note="Vector: pBel0BAC 11; Site 1: Hind III. Partially Hind III digested and size-selected S. mansoni cercarial DNA was ligated into Hind III digested pBel0BAC 11 vector and used to transform E. coli DH10B. The complete library contains 23808 clones from 4 independent sizing-ligation-transformations. Average insert size ranges from 70-127 kb and genome coverage is 7.9-fold."
ORIGIN	
Query Match	5.9%; Score 35; DB 28; Length 814;
Best Local Similarity	63.9%; Pred. No. 16;
Matches	53; Conservative 0; Mismatches 30; Indels 0; Gaps 0
CY	193 AAGTTACGCGCGCATTCAGACGCTTATTTCAATAGTTTCGAGAGGAAATGTG 252
DB	368 AAATGCACTTGGTCATACGATGATGATTAAGTCTCATGCAATTCGGCAGAAATCTGTA 329
CY	253 GTATTAACAGAGGTGACCTTAA 275
DB	328 GTATTAACAAAGGTGGCAATTA 306
RESULT 15	
CNS07K3H/c	814 bp DNA linear GSS 30-NOV-2001
LOCUS	T3 end of clone 010AD11 of library SmbAC1 from strain Puerto-Rican
DEFINITION	of Schistosoma mansoni, genomic survey sequence.
ACCESSION	AL614543
VERSION	AL614543.1 GI:16027767
KEYWORDS	GSS.
SOURCE	Schistosoma mansoni
ORGANISM	Schistosoma mansoni
REFERENCE	Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigoida; Schistosomatoidae; Schistosomatidae; Schistosoma.
AUTHORS	1 (bases 1 to 814) Le Paslier, M.-C., Pierce, R. J., Merlin, F., Hirai, H., Wu, W., Williams, D. L., Johnston, D., Loverde, P. T. and Le Paslier, D. Construction and characterization of a Schistosoma mansoni
TITLE	

```

JOURNAL      Bacterial artificial chromosome library
Genomics    65(2), 87-94 (2000)
MEDLINE      20247247
PubMed      10783255
REFERENCE    2 (bases 1 to 814)
AUTHORS
TITLE
JOURNAL
COMMENT
Genoscope.
Direct Submision
Submitted (05-OCr-2001) Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
Partially Hind III digested and size-selected S. mansoni cercarial
DNA was ligated into Hind III digested pBlotAC 11 vector and used
to transform E. coli DH10B. The complete library contains 23808
clones from 4 independent sizing-ligation-transformations. Average
insert size ranges from 70-127 kb and genome coverage is 7.9-fold.
FEATURES
source
1..814
/organism="Schistosoma mansoni"
/mol_type="genomic DNA"
/strain="Puerto-Rican"
/db_xref="taxon:6183"
/clone="010AD11"
/clone_11b="SmbAC1"
/note="end : T3"
ORIGIN
Query Match      5.9% Score 35; DB 29; Length 814;
Best Local Similarity 63.9%; Pred. No. 16;
Matches 53; Conservative 0; Mismatches 30; Indels 0; Gaps 0
Dy
193 AACGTTACGCGGCAGATTCAGACGTTATATTTTCAATAGATTTCGAGAGAGAAATGTG 252
Db
388 AATTCCAACTTGTCATCATCAGTACTATTAATGCTCTCATCGAATTCGGCCAGATACTGTA 329
Dy
253 GTAGTAAAGAGAGGTGACCTTAA 275
Db
328 GTATTACAAAGCGGTGGCAATTA 306

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 17:46:17 ; Search time 60 Seconds
(without alignments)
932.407 Million cell updates/sec

Title: US-09-613-486-15

Perfect score: 991
Sequence: 1 MELMSDNLNLTVDASSL.....GGVNTPVSNLRQLGRREVM 198

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1980s:*
3: geneseqp2000s:*
4: geneseqp2000s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	991	100.0	198	2	AAW73482 Grapevine
2	286	28.9	204	2	AAW72682 Sugar Bee
3	89.5	9.0	499	5	ABG32273 C. albica
4	86.5	8.7	328	5	ABB31963 Herbicida
5	86	8.7	295	3	AAAG39167 Arabidops
6	86	8.7	295	3	AAAG20735 Arabidops
7	86	8.7	360	3	AAAG20734 Arabidops
8	86	8.7	360	3	AAAG39166 Arabidops
9	86	8.7	360	5	ABB31962 Herbicida
10	86	8.7	366	3	AAAG39165 Arabidops
11	86	8.7	366	3	AAAG20733 Arabidops
12	86	8.7	447	4	AAAB6072 Putative
13	85	8.6	599	4	AAW17788 Phage abo
14	84.5	8.5	269	4	AAV97690 GRRAV-5 C
15	83.5	8.4	422	6	ABU26015 Protein e
16	83.5	8.3	591	6	ABU24731 Protein e
17	81.5	8.2	402	5	AAAB4800 Erysipelo
18	81.5	8.2	402	5	AAAB08869 Erysipelo
19	81.5	8.2	402	5	ABBO787 Erysipelo
20	81.5	8.2	626	5	AAV97782 Erysipelo
21	81.5	8.2	813	6	ABU19284 Protein e
22	80	8.1	555	5	ABBS4251 Streptoco
23	80	8.1	711	5	ABP29995 Streptoco
24	80	8.1	711	6	ABU46760 Protein e
25	80	8.1	729	5	ABP25672 Streptoco

26	79.5	8.0	770	2	AAW26584 Rat hemid
27	79	8.0	263	4	AAU49697 Propionib
28	79	8.0	263	6	ABM46216 Propionib
29	79	8.0	451	4	AAU35128 Enterococ
30	79	8.0	451	6	ABU29218 Protein e
31	79	8.0	477	2	AAW21643 Grapevine
32	79	8.0	478	6	ABU35792 Protein e
33	79	8.0	782	4	ABBS57835 Drosophila
34	78.5	7.9	208	5	ABBS4384 Lactococc
35	77	7.8	696	5	ABP73971 Candida a
36	77	7.8	865	6	ABU22392 Protein e
37	76.5	7.7	200	4	AAU30591 Novel hum
38	76	7.7	223	2	AAW73481 Grapevine
39	75.5	7.6	509	2	AAW47583 NADH oxid
40	75.5	7.6	606	2	AAV43219 E. thussio
41	75	7.6	358	5	ABBS1943 Herbicida
42	75	7.6	404	5	ABP25433 Streptoco
43	75	7.6	883	4	AAW51651 Synchoco
44	75	7.6	26926	4	AAU05396 Human tit
45	75	7.6	31267	6	ABG74786 Human RGS

ALIGNMENTS

RESULT 1
AAW73482
ID AAW73482 standard; protein, 198 AA.

AC AAW73482;
XX
XX
DT 27-MUG-2003 (revised)
DT 29-MAR-1999 (first entry)
XX
XX
DE Grapevine leafroll virus type 2 coat protein.
XX
XX
KM GRNAV-2; closterovirus; grape, tobacco; transgenic plant;
KM disease resistance; virus resistance; beet yellows virus; cristeza virus;
KM coat protein.
XX
XX
OS Grapevine leafroll virus.
OS
PN WO9853055-A1
XX
XX
PD 26-NOV-1998.
XX
XX
PF 20-MAY-1998; 98MO-US010313.
XX
XX
PR 20-MAY-1997; 97US-0047194P.
XX
XX
PA (CORR) CORNELL RES FOUND INC.
XX
XX
PI Zhu, H., Ling, K., Gonsalves D;
XX
XX
DR WPI. 1999-045307/04.
DR N-PSDB; AAV08870.
XX
XX
PT Grapevine leafroll virus (type 2) proteins and polypeptides - and
PT encoding DNA, useful e.g. to impart grapevine leafroll resistance to
PT grape and tobacco plants and detect grapevine leafroll virus.
XX
XX
PS Claim 12; Page 44-45; 151pp; English.
XX
XX
CC This is the amino acid sequence of a 22 kDa coat protein that is encoded
CC by open reading frame ORF6 (see AAV08870) of grapevine leafroll virus
CC type 2 (GRNAV-2) RNA (see AAV08874). The GRNAV-2 genome includes 9 open
CC reading frames (see AAV08864-72) for a polyprotein, an RNA-dependent RNA
CC polymerase, heat shock proteins, coat proteins and proteins of unknown
CC function (see AAW73476-84). These can be used to produce antibodies
CC useful for detecting GRNAV in samples e.g. by ELISA (claimed). The
CC nucleic acid molecules can be used to produce probes and primers for such
CC detection, and to transform host cells (especially Agrobacterium vitis,
CC Agrobacterium tumefaciens, grape, citrus, beet or tobacco cells) and

CC produce transgenic plants (claimed). They can be used to impart GLRAV-2
 CC resistance to Vitis scion or rootstock cultivars or Nicotiana cultivars
 CC (claimed). Because extensive similarity exists between hsp70-related
 CC sequence regions of GLRAV-2 and other closteroviruses, the DNA may also
 CC be used to impart beet yellows virus resistance to beet cultivars or
 CC tristea virus resistance to citrus scion cultivar/rootstock cultivars
 CC (claimed). (Updated on 27-Aug-2003 to correct OS field.)

CC Sequence 198 AA;

Query Match 100.0%; Score 991; DB 2; Length 198;
 Best Local Similarity 100.0%; Pred. No. 2.9e-100;
 Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELMSDSNLSNLTVDASSLNGVDKLLSAEVEKMLVQKAPNEGIEVVFGLLYALAAAR 60
 DB 1 MELMSDSNLSNLTVDASSLNGVDKLLSAEVEKMLVQKAPNEGIEVVFGLLYALAAAR 60
 QY 61 TTSPKVOQADSDVIFSNFSGERNVVTBGLDKVLDGCAPLRTPTNKLRTFGRTPEAYV 120
 DB 61 TTSPKVOQADSDVIFSNFSGERNVVTBGLDKVLDGCAPLRTPTNKLRTFGRTPEAYV 120
 QY 121 DFCIAYKHKLPOLNAAAEELGIPADSYLAADFLGTCPKLSLQSRKMPASMYALKTEGG 180
 DB 121 DFCIAYKHKLPOLNAAAEELGIPADSYLAADFLGTCPKLSLQSRKMPASMYALKTEGG 180
 QY 181 VVNTPVSNLRQLGRREV 198
 DB 181 VVNTPVSNLRQLGRREV 198

RESULT 2

AAR72682 standard; protein; 204 AA.

AC AAR72682;
 DT 16-OCT-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 02-NOV-1995 (first entry)

XX Sugar beet yellows virus capsid protein.

XX Sugar beet yellows virus capsid protein; transgenic plant.

XX Beet yellows virus.

XX RU2017820-C1.

XX 15-AUG-1994.

XX 27-JUN-1991; 91SU-04950054.

XX 27-JUN-1991; 91SU-04950054.

XX (IMMUNO) IMMUNOBIOTECH INST.

XX (BIOT=) BIOYRCHN INST CO LTD.

XX Arganovskii AA, Bolko VP, Karasev AV;

XX MPI; 1995-113715/15.

XX N-PSDB; AAQ87653.

XX Sugar beet yellows virus cDNA fragment encoding capsid protein - useful
 for production of virus-resistant transgenic plants.

XX Claim 1; Col 7-10; 5pp; Russian.

CC The amino acid sequence of the novel sugar beet yellows virus (SBYV)
 CC capsid protein. The protein has mol. wt 22.2 kD. The corresponding gene
 CC was obtained from reverse transcribed RNA isolated from purified SBYV.
 CC The fragment is useful for the production of virus resistant transgenic
 CC plants by genetic engineering methods. (Updated on 25-MAR-2003 to correct

CC PN field.) (Updated on 25-MAR-2003 to correct PF field.) (Updated on 16-
 CC OCT-2003 to standardise OS field)

CC Sequence 204 AA;

Query Match 28.9%; Score 286; DB 2; Length 204;
 Best Local Similarity 37.6%; Pred. No. 8.5e-23;
 Matches 71; Conservative 29; Mismatches 87; Indels 2; Gaps 2;

QY 7 SNLSNLTVDASSLNGVDKLLSAEVEKMLVQKAPNEGIEVVFGLLYALAAAR 66
 DB 11 AFFENVSLADQCLHGECDDKLRNFEECLKLKGPEENLGLALGLCYSCATIGTSNKV 70
 QY 67 QPADSDVIFSNF-GERNVVTEBGLDKVLDGCAPLRTPTNKLRTFGRTPEAYVDFCIA 125
 DB 71 NQPTSTTKASFGGKSLVTHGELNSFLSQKLSKPKKLCFCRTPOKOTISLKE 130
 QY 126 YHKLPOLNAAAEELGIPADSYLAADFLGTCPKLSLQSRKMPASMYALKTEGGVNT 185
 DB 131 YRGKLPPIAARHGLPADHYLAADFISTSTELDLQSRLLARENATHTERS-SESP 189
 QY 186 VSNLRQLGR 194
 DB 190 VSNLRQLGR 198

RESULT 3

ABG93273 standard; protein; 499 AA.

AC ABG93273;

DT 21-NOV-2002 (first entry)

XX C. albicans BAX-associated protein fragment SEQ ID 504.

XX Bax; Bax-resistance; cytosolic; fungicide; immunosuppressive; virucide;

XX vasotropic; vaccine; gene therapy; proliferative disorder; cancer;

XX apoptosis; fungal; yeast; infection; autoimmune disease; ischaemia;

XX neurodegeneration; cell death.

XX Candida albicans.

XX WO200264766-A2.

XX 22-AUG-2002.

XX 21-DEC-2001; 2001WO-EP015398.

XX 22-DEC-2000; 2000EP-00870318.

XX 04-JAN-2001; 2001EP-00870002.

XX 09-JAN-2001; 2001EP-00870003.

XX (JANC) JANSSEN PHARM NV.

XX Contreras RH, Eberhardt I, Luyten WHML, Reekmans RJ;

XX MPI; 2002-667002/71.

XX N-PSDB; ABQ76539.

XX New isolated nucleic acid representing a synthetic BAX-gene, useful as
 medicament for treating, preventing and/or alleviating yeast or fungal
 PT infections or proliferative disorders, or for preventing apoptosis in
 PT certain diseases.

XX Claim 36; Fig 2; 344pp; English.

CC This invention describes a novel nucleic acid representing a synthetic
 CC Bax gene. The Bax gene of the invention is useful for identifying Bax-
 CC resistant yeast or fungi, identifying, or obtaining and identifying
 CC Candida spp. sequences that are differentially expressed in a pathway
 CC eventually leading to programmed cell death or identifying inhibitors or
 CC inhibitor sequences of Bax-induced cell death. The products of the

CC invention have cytostatic, fungicide; immunosuppressive, virucide and
 CC vasotropic activity and can be used in vaccines or for gene therapy. The
 CC isolated nucleic acids, polypeptides, pharmaceutical compositions,
 CC antitense molecules and antibodies are useful as medicaments or in
 CC preparing a medicament for treating, preventing and/or alleviating
 CC diseases associated with yeast or fungi or proliferative disorders, such
 CC as cancer, or for preventing apoptosis in certain diseases. The compounds
 CC or polypeptides, or the genetically modified organism are useful for
 CC preparing a medicament for modifying the endogenous flora of humans and
 CC other mammals. The vaccine is useful for immunising against yeast or
 CC fungal infections. Apoptosis-related diseases include autoimmune diseases,
 CC ischaemia, diseases related with viral infections or neurodegenerations.
 CC This sequence represents a polypeptide associated with the Bax gene
 CC described in the disclosure of the invention

XX Sequence 499 AA;

Query Match 9.0%; Score 89.5; DB 5; Length 499;
 Best Local Similarity 25.1%; Pred. No. 1.2;
 Matches 53; Conservative 39; Mismatches 76; Indels 43; Gaps 12;

QY 3 LMSDSNLSNLTVPASS-LNGVDKLLSA-EVEKTVQKAPNEGIEVFGLLVALAR 60
 DB 173 LMSWKLGPVLATGTTVLTAEFTPLSALYLSQLLEAGMPGVNIVSGFATGCAAI 232
 QY 61 TTSPKVRADSDVIFSNFGERNVV---TEGDLKKV---LDGCAPLTRFTNKLRTFGRT 114
 DB 233 AKHKIKR---VAFTGSTATGKIIMLAESNLKVTLELGGKSPRIIVND---ADLDKT 286
 QY 115 FTEAYV-----DFCTA-----YKHKLPQMAAE---LGTF-AEDSYLADFL 153
 DB 287 IONLIVIFPNSGVCCAGSRLLIQSGVYDQVVEKFAESYKGVNPFDEDTFMA--- 343
 QY 154 GTCFKLSELOOSRKMFPASWYALKTEGGVNT 184
 DB 344 ---QVSDVQLS-KILKTVESGKSGCATVVT 369

RESULT 4

ABB91963 standard; protein; 328 AA.

XX ABB91963;
 XX 31-MAY-2002 (first entry)
 XX Herbicidally active polypeptide SEQ ID NO 1174.
 XX Herbicidally plant; agriculture; herbicide.
 XX Arabidopsis thaliana.
 XX WO200210210-A2.
 XX 07-FEB-2002.
 XX 28-AUG-2001; 2001WO-EP009892.
 XX 28-AUG-2001; 2001WO-EP009892.
 XX (FARB) BAYER AG.
 XX Tietjen K, Weidner M;
 XX WPT 2002-269010/31.
 XX Identifying plant target proteins for herbicidally active compounds,
 XX comprising aligning and comparing nucleic acid or amino acid sequences
 XX from plant with nucleic acid or amino acid sequences from non-plant
 XX organisms.
 XX Claim 5; SEQ ID NO 1174; 261pp + Sequence Listing; English.

CC The invention relates to identifying target proteins (ABB90790-ABB94016)
 CC for herbicidally active compounds, comprising aligning and comparing
 CC nucleic acid or amino acid sequences from plant with nucleic acid or
 CC amino acid sequences from non-plant organisms using suitable search
 CC parameters, where plant sequences having an E-value greater by a factor
 CC of 3 than the E-value of most similar non-plant sequences are selected.
 CC The polypeptides or nucleic acids encoding them are useful for
 CC identifying modulators. The identified modulators are useful as
 CC herbicides

XX Sequence 328 AA;

Query Match 8.7%; Score 86.5; DB 5; Length 328;
 Best Local Similarity 25.8%; Pred. No. 1.4;
 Matches 49; Conservative 18; Mismatches 86; Indels 37; Gaps 10;

QY 14 ITDASSLNGV-----DKLLSAEVEKTVQKAPNEGIEVFGLLVALARTSP 64
 DB 105 LSDQILTVGVPASAGIVGDKKAMEIINNAFVVSAGEND-----FILNTYDIFSRLLEY 159
 QY 65 KVQRADSDVIFSN-----SFGERNVVTTEGDLKKVLDGCAPLTRFTNKLRTFGRT 115
 DB 160 PRISGVQDFILKRLNFVRELYSICVRNVLV---GGLPRM---GCLPI-HMTAKFRNIFRFC 214
 QY 116 TEAYVDFCIAYKHKLPQMAAEICIPAESYLADEFCTGPKXSELQOSKMFASWAL 175
 DB 215 LEHNKRDVLTNEXKQKLLPQIEASLPG-SKFLYADVYN---PMMEMTONPSK-----YGF 266
 QY 176 K-TEGGVNT 184
 DB 267 KETKRGCCGT 276

RESULT 5

AAG39167 standard; protein; 295 AA.

XX AAG39167;
 XX 18-OCT-2000 (first entry)
 XX Arabidopsis thaliana protein fragment SEQ ID NO: 48421.
 XX Protein identification; signal transduction pathway; metabolic pathway;
 XX hybridisation assay; genetic mapping; gene expression control; promoter;
 XX termination sequence.
 XX Arabidopsis thaliana.
 XX EP1033405-A2.
 XX 06-SEP-2000.
 XX 25-FEB-2000; 2000EP-00301439.
 XX 25-FEB-1999; 99US-0121825P.
 XX 05-MAR-1999; 99US-0123180P.
 XX 09-MAR-1999; 99US-0123548P.
 XX 23-MAR-1999; 99US-0125788P.
 XX 25-MAR-1999; 99US-0126264P.
 XX 29-MAR-1999; 99US-0126785P.
 XX 01-APR-1999; 99US-0127462P.
 XX 06-APR-1999; 99US-0128334P.
 XX 08-APR-1999; 99US-0128714P.
 XX 16-APR-1999; 99US-0129845P.
 XX 19-APR-1999; 99US-0130077P.
 XX 21-APR-1999; 99US-0130449P.
 XX 23-APR-1999; 99US-0130510P.
 XX 23-APR-1999; 99US-0130891P.
 XX 28-APR-1999; 99US-0131449P.
 XX 30-APR-1999; 99US-0132048P.
 XX 30-APR-1999; 99US-0132407P.
 XX 04-MAY-1999; 99US-0132484P.

PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132488P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135253P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136332P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139453P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139489P.
PR 18-JUN-1999; 99US-0139484P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140351P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0141843P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 21-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.

PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145122P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148177P.
PR 11-AUG-1999; 99US-0148341P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 27-AUG-1999; 99US-0151303P.
PR 30-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 01-SEP-1999; 99US-0152363P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153158P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154779P.
PR 20-SEP-1999; 99US-0155139P.
PR 22-SEP-1999; 99US-0155486P.
PR 23-SEP-1999; 99US-0155659P.
PR 24-SEP-1999; 99US-0156458P.
PR 28-SEP-1999; 99US-0156596P.
PR 29-SEP-1999; 99US-0157117P.
PR 04-OCT-1999; 99US-0157753P.
PR 05-OCT-1999; 99US-0157865P.
PR 06-OCT-1999; 99US-0158029P.
PR 07-OCT-1999; 99US-0158232P.
PR 08-OCT-1999; 99US-0158369P.
PR 12-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 13-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159684P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.

PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 22-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 25-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 26-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query March 8.7%; Score 86; DB 3; Length 295;
Best Local Similarity 26.5%; Pred. No. 1.3;

Matches 49; Conservative 23; Mismatches 93; Indels 20; Gaps 9;

QY 7 SNLSNLTVDASSINGV--DKKLISAEVERKMLYOKGAPNNGT---EVVFGLLLYLAAR 60
DB 72 SEQENMFKSYIARLKGIVGDKKMEIINNAFVWVSAGPNDFINYYEIPSRRLYEYFISG 121
QY 61 TTSPKVGRAQSDVFFNSFGERNVVTGDLKKVLDGCAPLTFRTNKLRTFGRTTEAYV 120
DB 132 YQDPIILKLENFVRELKSLGVANLV--GGLPFW--GCLPI-HMTAKFRNIFFCLEHNN 186
QY 121 DFCIAYRHKLPGQINAAAEELGIPADSYLADFLGTCKLSELOQSRMFASMTALK-TEG 179
DB 187 KDSVLYNEKIQNLPQTEASLPG-SKFLYADVVA--EMMEMIQVPSK-----YGFKETKR 238
QY 180 GVWNT 184
DB 239 GCCGT 243

RESULT 6

AAG20735
ID AAG20735 standard; protein; 295 AA.

XX AAG20735;

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 23037.

XX Protein identification; signal transduction pathway; metabolic pathway;
XX Hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.

PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134321P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 18-MAY-1999; 99US-0134941P.
PR 19-MAY-1999; 99US-0135124P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136393P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0138111P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-01442390P.
PR 08-JUL-1999; 99US-01442803P.
PR 09-JUL-1999; 99US-01442820P.
PR 12-JUL-1999; 99US-01442977P.
PR 13-JUL-1999; 99US-0144342P.
PR 14-JUL-1999; 99US-01443624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 18-JUL-1999; 99US-0144325P.
PR 18-JUL-1999; 99US-0144331P.
PR 18-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.

PR 20-JUL-1999; 99US-014484P.
 PR 21-JUL-1999; 99US-014481P.
 PR 21-JUL-1999; 99US-0145086P.
 PR 21-JUL-1999; 99US-0145088P.
 PR 22-JUL-1999; 99US-0145085P.
 PR 22-JUL-1999; 99US-0145087P.
 PR 22-JUL-1999; 99US-0145089P.
 PR 22-JUL-1999; 99US-0145192P.
 PR 23-JUL-1999; 99US-0145145P.
 PR 23-JUL-1999; 99US-0145218P.
 PR 23-JUL-1999; 99US-0145224P.
 PR 25-JUL-1999; 99US-0145276P.
 PR 27-JUL-1999; 99US-0145913P.
 PR 27-JUL-1999; 99US-0145918P.
 PR 27-JUL-1999; 99US-0145919P.
 PR 28-JUL-1999; 99US-0145951P.
 PR 02-AUG-1999; 99US-0146386P.
 PR 02-AUG-1999; 99US-0146388P.
 PR 02-AUG-1999; 99US-0146389P.
 PR 03-AUG-1999; 99US-0147038P.
 PR 04-AUG-1999; 99US-0147204P.
 PR 04-AUG-1999; 99US-0147302P.
 PR 05-AUG-1999; 99US-0147192P.
 PR 05-AUG-1999; 99US-0147260P.
 PR 06-AUG-1999; 99US-0147303P.
 PR 06-AUG-1999; 99US-0147416P.
 PR 09-AUG-1999; 99US-0147493P.
 PR 09-AUG-1999; 99US-0147935P.
 PR 10-AUG-1999; 99US-0148171P.
 PR 11-AUG-1999; 99US-0148319P.
 PR 12-AUG-1999; 99US-0148341P.
 PR 13-AUG-1999; 99US-0148565P.
 PR 13-AUG-1999; 99US-0148684P.
 PR 16-AUG-1999; 99US-0149368P.
 PR 17-AUG-1999; 99US-0149175P.
 PR 18-AUG-1999; 99US-0149426P.
 PR 20-AUG-1999; 99US-0149722P.
 PR 20-AUG-1999; 99US-0149723P.
 PR 20-AUG-1999; 99US-0149929P.
 PR 23-AUG-1999; 99US-0149902P.
 PR 23-AUG-1999; 99US-0149930P.
 PR 23-AUG-1999; 99US-0150566P.
 PR 26-AUG-1999; 99US-0150884P.
 PR 27-AUG-1999; 99US-0151065P.
 PR 27-AUG-1999; 99US-0151066P.
 PR 27-AUG-1999; 99US-0151080P.
 PR 30-AUG-1999; 99US-0151303P.
 PR 31-AUG-1999; 99US-0151438P.
 PR 01-SEP-1999; 99US-0151930P.
 PR 07-SEP-1999; 99US-0152363P.
 PR 10-SEP-1999; 99US-0153070P.
 PR 13-SEP-1999; 99US-0153758P.
 PR 15-SEP-1999; 99US-0154018P.
 PR 16-SEP-1999; 99US-0154039P.
 PR 20-SEP-1999; 99US-0154779P.
 PR 22-SEP-1999; 99US-0155139P.
 PR 23-SEP-1999; 99US-0155148P.
 PR 24-SEP-1999; 99US-0155659P.
 PR 28-SEP-1999; 99US-0156458P.
 PR 29-SEP-1999; 99US-0156596P.
 PR 04-OCT-1999; 99US-0157117P.
 PR 05-OCT-1999; 99US-0157753P.
 PR 06-OCT-1999; 99US-0157865P.
 PR 07-OCT-1999; 99US-0158028P.
 PR 08-OCT-1999; 99US-0158232P.
 PR 12-OCT-1999; 99US-0158369P.
 PR 13-OCT-1999; 99US-0159293P.
 PR 13-OCT-1999; 99US-0159294P.
 PR 13-OCT-1999; 99US-0159295P.
 PR 14-OCT-1999; 99US-0159329P.
 PR 14-OCT-1999; 99US-0159330P.
 PR 14-OCT-1999; 99US-0159331P.
 PR 14-OCT-1999; 99US-0159637P.

PR 14-OCT-1999; 99US-0159638P.
 PR 18-OCT-1999; 99US-0159584P.
 PR 21-OCT-1999; 99US-0160741P.
 PR 21-OCT-1999; 99US-0160767P.
 PR 21-OCT-1999; 99US-0160768P.
 PR 21-OCT-1999; 99US-0160770P.
 PR 21-OCT-1999; 99US-0160814P.
 PR 21-OCT-1999; 99US-0160815P.
 PR 22-OCT-1999; 99US-0160980P.
 PR 22-OCT-1999; 99US-0160981P.
 PR 22-OCT-1999; 99US-0160989P.
 PR 22-OCT-1999; 99US-0161404P.
 PR 25-OCT-1999; 99US-0161405P.
 PR 25-OCT-1999; 99US-0161406P.
 PR 25-OCT-1999; 99US-0161359P.
 PR 26-OCT-1999; 99US-0161360P.
 PR 26-OCT-1999; 99US-0161361P.
 PR 26-OCT-1999; 99US-0161920P.
 PR 28-OCT-1999; 99US-0161920P.
 PR 28-OCT-1999; 99US-0161922P.
 PR 28-OCT-1999; 99US-0161932P.
 PR 29-OCT-1999; 99US-0162142P.

Query Match 8.7%; Score 86; DB 3; Length 295;
 Best Local Similarity 26.5%; Pred. No. 1.3;
 Matches 49; Conservative 23; Mismatches 93; Indels 20; Gaps 9;

QY 7 SNLSNLYITDASSLNGV--DKGLSAVEKRLVQKCAPNEGI---EIVFGLLYALAA 60
 DB 72 SEQPMFPRXYIARLKGIVGDKKAMEIINNAFVVSAGNDFILNYEIPSRRLYPPISG 131
 QY 61 TSPKVPQADSPVIFSNFGERNVYVTEGDKKYLDCGAPLTFRTNKLRTGRTTAYV 120
 DB 132 YQDFLKLLENVRLYSLGRNVYV--GGLPVM--GGLPI--HMTAKFRNIFRCLSEHN 186
 QY 121 DPCIAVKKRLPQNAALGLIPADSVYADPLGTCPKLSLQSRKKFASMYALK--TEG 179
 DB 187 KDSVLYNEKLNILPQIBASLPG--SKFLYADVYN--PMMENIQNPSK----YGRKETKR 238
 QY 180 GYVNT 184
 DB 239 GCCGT 243

RESULT 7
 AAG20734
 ID AAG20734 standard; protein; 360 AA.
 XX
 AC AAG20734;
 XX
 DT 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 23036.
 XX
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KM hybridisation assay; genetic mapping; gene expression control; promoter;
 XX termination sequence.
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-00301439.
 XX
 PR 25-FEB-1999; 99US-0121825P.
 PR 05-MAR-1999; 99US-0123180P.
 PR 09-MAR-1999; 99US-0123548P.
 PR 23-MAR-1999; 99US-0125788P.
 PR 25-MAR-1999; 99US-0126264P.
 PR 29-MAR-1999; 99US-0126785P.
 PR 01-APR-1999; 99US-0127462P.
 PR 06-APR-1999; 99US-0128234P.

PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0128845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 18-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 24-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140822P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142800P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.

PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145086P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147446P.
PR 09-AUG-1999; 99US-0147449P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 13-AUG-1999; 99US-0149368P.
PR 16-AUG-1999; 99US-0149175P.
PR 17-AUG-1999; 99US-0149426P.
PR 18-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 01-SEP-1999; 99US-0152362P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154799P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0156589P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157751P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158322P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.

PR 13-OCT-1999; 99US-0159295P.
 PR 14-OCT-1999; 99US-0159329P.
 PR 14-OCT-1999; 99US-0159330P.
 PR 14-OCT-1999; 99US-0159331P.
 PR 14-OCT-1999; 99US-0159637P.
 PR 14-OCT-1999; 99US-0159638P.
 PR 18-OCT-1999; 99US-0159584P.
 PR 21-OCT-1999; 99US-0160741P.
 PR 21-OCT-1999; 99US-0160747P.
 PR 21-OCT-1999; 99US-0160768P.
 PR 21-OCT-1999; 99US-0160770P.
 PR 21-OCT-1999; 99US-0160814P.
 PR 21-OCT-1999; 99US-0160815P.
 PR 22-OCT-1999; 99US-0160980P.
 PR 22-OCT-1999; 99US-0160981P.
 PR 22-OCT-1999; 99US-0160989P.
 PR 22-OCT-1999; 99US-0161404P.
 PR 22-OCT-1999; 99US-0161405P.
 PR 25-OCT-1999; 99US-0161406P.
 PR 26-OCT-1999; 99US-0161559P.
 PR 26-OCT-1999; 99US-0161360P.
 PR 26-OCT-1999; 99US-0161361P.
 PR 28-OCT-1999; 99US-0161820P.
 PR 28-OCT-1999; 99US-0161822P.
 PR 28-OCT-1999; 99US-0161933P.
 PR 29-OCT-1999; 99US-0162142P.

Query Match 8.7%; Score 86; DB 3; Length 360;
 Best Local Similarity 26.5%; Pred. No. 18;
 Matches 49; Conservative 23; Mismatches 93; Indels 20; Gaps 9;

QY 7 SNLSNIVITDASSINCV--DKLLSAEVEKMLVQKGAENEGL---EYVFGLLLYALAAR 60
 DB 137 SEOPNPFKYLARKLKGIVGDKKAMEIINNAFVVSAGNDPLNYEILPSRLLEYPFISG 196
 QY 61 TTSFKVQADSDVIEFNSFGERNVVTEGDLKKVLDGCAPLTRFTNKLRTGRTETAYV 120
 DB 197 YQDFILKRLNFVRELYSIGVANLV--GGLFPM--GGLPI--HMTAKFRNIFRCLBHN 251
 QY 121 DFCIAKHLPLQNAALAEIGTIAEDSYLAADPLGTCPLSELQSRKMFAMYLK--TEG 179
 DB 252 KQSVLNEKQNLPLQIEASLPG--SKFLYADVYN--PMMEMIQNPSK----YGFKEKTKR 303
 QY 180 GVYNT 184
 DB 304 GCCGT 308

RESULT 8
 AAG39166
 ID AAG39166 standard; protein; 360 AA.

XX AAG39166;
 DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 48420.

KW Protein identification; signal transduction pathway; metabolic pathway;
 hybridisation assay; genetic mapping; gene expression control; promoter;
 termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.
 PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.
 PR 25-MAR-1999; 99US-0126264P.
 PR 29-MAR-1999; 99US-0126785P.
 PR 01-APR-1999; 99US-0127463P.
 PR 06-APR-1999; 99US-0128234P.
 PR 08-APR-1999; 99US-0128714P.
 PR 16-APR-1999; 99US-0128845P.
 PR 19-APR-1999; 99US-0130077P.
 PR 21-APR-1999; 99US-0130449P.
 PR 23-APR-1999; 99US-0130510P.
 PR 23-APR-1999; 99US-0130891P.
 PR 28-APR-1999; 99US-0131449P.
 PR 30-APR-1999; 99US-0132048P.
 PR 04-MAY-1999; 99US-0132407P.
 PR 05-MAY-1999; 99US-0132484P.
 PR 06-MAY-1999; 99US-0132486P.
 PR 06-MAY-1999; 99US-0132487P.
 PR 07-MAY-1999; 99US-0132863P.
 PR 11-MAY-1999; 99US-0134256P.
 PR 14-MAY-1999; 99US-0134218P.
 PR 14-MAY-1999; 99US-0134219P.
 PR 14-MAY-1999; 99US-0134321P.
 PR 14-MAY-1999; 99US-0134370P.
 PR 18-MAY-1999; 99US-0134768P.
 PR 19-MAY-1999; 99US-0134941P.
 PR 20-MAY-1999; 99US-0135124P.
 PR 21-MAY-1999; 99US-0135353P.
 PR 24-MAY-1999; 99US-0135629P.
 PR 25-MAY-1999; 99US-0136021P.
 PR 27-MAY-1999; 99US-0136392P.
 PR 28-MAY-1999; 99US-0136782P.
 PR 01-JUN-1999; 99US-0137222P.
 PR 03-JUN-1999; 99US-0137528P.
 PR 04-JUN-1999; 99US-0137502P.
 PR 07-JUN-1999; 99US-0137724P.
 PR 08-JUN-1999; 99US-0138049P.
 PR 10-JUN-1999; 99US-0138541P.
 PR 10-JUN-1999; 99US-0138847P.
 PR 14-JUN-1999; 99US-0139119P.
 PR 16-JUN-1999; 99US-0139452P.
 PR 16-JUN-1999; 99US-0139453P.
 PR 17-JUN-1999; 99US-0139492P.
 PR 18-JUN-1999; 99US-0139454P.
 PR 18-JUN-1999; 99US-0139455P.
 PR 18-JUN-1999; 99US-0139456P.
 PR 18-JUN-1999; 99US-0139457P.
 PR 18-JUN-1999; 99US-0139458P.
 PR 18-JUN-1999; 99US-0139459P.
 PR 18-JUN-1999; 99US-0139460P.
 PR 18-JUN-1999; 99US-0139461P.
 PR 18-JUN-1999; 99US-0139462P.
 PR 18-JUN-1999; 99US-0139463P.
 PR 18-JUN-1999; 99US-0139780P.
 PR 18-JUN-1999; 99US-0139781P.
 PR 18-JUN-1999; 99US-0139782P.
 PR 21-JUN-1999; 99US-0139817P.
 PR 22-JUN-1999; 99US-0139899P.
 PR 23-JUN-1999; 99US-0140353P.
 PR 23-JUN-1999; 99US-0140354P.
 PR 24-JUN-1999; 99US-0140659P.
 PR 28-JUN-1999; 99US-0140823P.
 PR 29-JUN-1999; 99US-0140991P.
 PR 30-JUN-1999; 99US-0141287P.
 PR 01-JUL-1999; 99US-0141842P.
 PR 01-JUL-1999; 99US-0142154P.
 PR 02-JUL-1999; 99US-0142055P.
 PR 06-JUL-1999; 99US-0142390P.
 PR 08-JUL-1999; 99US-0142803P.
 PR 09-JUL-1999; 99US-0142920P.
 PR 12-JUL-1999; 99US-0142977P.
 PR 13-JUL-1999; 99US-0143542P.
 PR 14-JUL-1999; 99US-0143624P.
 PR 15-JUL-1999; 99US-0144005P.

PR 16-JUL-1999; 99US-0144085P.
 PR 16-JUL-1999; 99US-0144086P.
 PR 19-JUL-1999; 99US-0144335P.
 PR 19-JUL-1999; 99US-0144331P.
 PR 19-JUL-1999; 99US-0144332P.
 PR 19-JUL-1999; 99US-0144333P.
 PR 19-JUL-1999; 99US-0144334P.
 PR 19-JUL-1999; 99US-0144335P.
 PR 20-JUL-1999; 99US-0144332P.
 PR 20-JUL-1999; 99US-0144884P.
 PR 21-JUL-1999; 99US-0144814P.
 PR 21-JUL-1999; 99US-0145086P.
 PR 21-JUL-1999; 99US-0145088P.
 PR 22-JUL-1999; 99US-0145085P.
 PR 22-JUL-1999; 99US-0145087P.
 PR 22-JUL-1999; 99US-0145089P.
 PR 22-JUL-1999; 99US-0145152P.
 PR 23-JUL-1999; 99US-0145145P.
 PR 23-JUL-1999; 99US-0145218P.
 PR 23-JUL-1999; 99US-0145224P.
 PR 26-JUL-1999; 99US-0145276P.
 PR 27-JUL-1999; 99US-0145913P.
 PR 27-JUL-1999; 99US-0145918P.
 PR 27-JUL-1999; 99US-0145919P.
 PR 28-JUL-1999; 99US-0145951P.
 PR 02-AUG-1999; 99US-0146386P.
 PR 02-AUG-1999; 99US-0146388P.
 PR 02-AUG-1999; 99US-0146389P.
 PR 03-AUG-1999; 99US-0147038P.
 PR 04-AUG-1999; 99US-0147204P.
 PR 04-AUG-1999; 99US-0147302P.
 PR 05-AUG-1999; 99US-0147192P.
 PR 05-AUG-1999; 99US-0147260P.
 PR 06-AUG-1999; 99US-0147303P.
 PR 06-AUG-1999; 99US-0147416P.
 PR 09-AUG-1999; 99US-0147493P.
 PR 09-AUG-1999; 99US-0147935P.
 PR 10-AUG-1999; 99US-0148171P.
 PR 11-AUG-1999; 99US-0148319P.
 PR 12-AUG-1999; 99US-0148341P.
 PR 13-AUG-1999; 99US-0148565P.
 PR 13-AUG-1999; 99US-0148684P.
 PR 16-AUG-1999; 99US-0149368P.
 PR 17-AUG-1999; 99US-0149175P.
 PR 18-AUG-1999; 99US-0149426P.
 PR 20-AUG-1999; 99US-0149722P.
 PR 20-AUG-1999; 99US-0149723P.
 PR 20-AUG-1999; 99US-0149923P.
 PR 23-AUG-1999; 99US-0149902P.
 PR 23-AUG-1999; 99US-0149930P.
 PR 25-AUG-1999; 99US-0150566P.
 PR 26-AUG-1999; 99US-0150884P.
 PR 27-AUG-1999; 99US-0151065P.
 PR 27-AUG-1999; 99US-0151066P.
 PR 27-AUG-1999; 99US-0151080P.
 PR 30-AUG-1999; 99US-0151303P.
 PR 31-AUG-1999; 99US-0151438P.
 PR 01-SEP-1999; 99US-0151930P.
 PR 07-SEP-1999; 99US-0152363P.
 PR 10-SEP-1999; 99US-0153070P.
 PR 13-SEP-1999; 99US-0153758P.
 PR 15-SEP-1999; 99US-0154018P.
 PR 16-SEP-1999; 99US-0154039P.
 PR 20-SEP-1999; 99US-0154779P.
 PR 22-SEP-1999; 99US-0155139P.
 PR 23-SEP-1999; 99US-0155486P.
 PR 24-SEP-1999; 99US-0155659P.
 PR 28-SEP-1999; 99US-0156458P.
 PR 29-SEP-1999; 99US-0156596P.
 PR 04-OCT-1999; 99US-0157117P.
 PR 05-OCT-1999; 99US-0157753P.
 PR 06-OCT-1999; 99US-0157865P.

PR 07-OCT-1999; 99US-0158029P.
 PR 08-OCT-1999; 99US-0158232P.
 PR 12-OCT-1999; 99US-0158346P.
 PR 13-OCT-1999; 99US-0158293P.
 PR 13-OCT-1999; 99US-0158294P.
 PR 13-OCT-1999; 99US-0158295P.
 PR 14-OCT-1999; 99US-0158329P.
 PR 14-OCT-1999; 99US-0158330P.
 PR 14-OCT-1999; 99US-0158331P.
 PR 14-OCT-1999; 99US-0158337P.
 PR 14-OCT-1999; 99US-0158638P.
 PR 18-OCT-1999; 99US-0158584P.
 PR 21-OCT-1999; 99US-0160741P.
 PR 21-OCT-1999; 99US-0160767P.
 PR 21-OCT-1999; 99US-0160768P.
 PR 21-OCT-1999; 99US-0160770P.
 PR 21-OCT-1999; 99US-0160814P.
 PR 21-OCT-1999; 99US-0160815P.
 PR 22-OCT-1999; 99US-0160981P.
 PR 22-OCT-1999; 99US-0160981P.
 PR 22-OCT-1999; 99US-0160989P.
 PR 25-OCT-1999; 99US-0161404P.
 PR 25-OCT-1999; 99US-0161405P.
 PR 25-OCT-1999; 99US-0161406P.
 PR 25-OCT-1999; 99US-0161359P.
 PR 26-OCT-1999; 99US-0161360P.
 PR 26-OCT-1999; 99US-0161361P.
 PR 26-OCT-1999; 99US-0161920P.
 PR 28-OCT-1999; 99US-0161922P.
 PR 28-OCT-1999; 99US-0161993P.
 PR 29-OCT-1999; 99US-0162142P.

Query Match 8.7%; Score 86; DB 3; Length 360;
 Best Local Similarity 26.5%; Pred. No. 1.8;
 Matches 49; Conservative 23; Mismatches 93; Indels 20; Gaps 9;

QY 7 SNLSNVLITDASSLGV--DKLLSAEYKMLVQKQAPNEG---EYVFGLLTYALAA 60
 DB 137 SEQPNMFYSYIRKLGIYGDKKAMEIINNAFVVSAGRDFLNTYVEIPSRLETFPFISG 196
 QY 61 ITSPKVRADSDVIFSNSEFGERNVVTEGDLKVLGCAPLTRFTNKLRTFGRTTEAV 120
 DB 197 YQDFIKRLENFVRELYSIGVNVLY--GGLPPM--GCLPI-HMTAKFRNIFRFLIEHNN 251
 QY 121 DFCIAVKHLPOLNAAELGIPAEBSYLAADPLGTCPEKLSHQSRKMFASVATK-TEG 179
 DB 252 KQSVLYNEKLNLLPOIEASLPG-SKFLYADVYN--PVMEMIQNPSK-----YGRKEIKR 303

QY 180 GVNT 184
 DB 304 GCCGT 308

RESULT 9
 ABB91962
 ID ABB91962 standard; protein; 360 AA.
 XX AC ABB91962;
 DT 31-MAY-2002 (first entry)
 XX DE Herbicidally active polypeptide SEQ ID NO 1173.
 XX KM Herbicidal; plant; agriculture; herbicide.
 XX OS Arabidopsis thaliana.
 XX EN W0200210210-A2.
 XX PD 07-FEB-2002.
 XX PF 28-AUG-2001; 2001MO-EP009892.

PR 28-AUG-2001; 2001WO-EP009892.
XX (FARB) BAYER AG.
XX Tievjen K, Weidler M;
PI WPI: 2002-269010/31.
XX
XX Identifying plant target proteins for herbicidally active compounds,
PR comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PR organisms.
XX
XX Claim 5; SEQ ID NO 1173; 261pp + Sequence listing; English.
XX
XX The invention relates to identifying target proteins (ABR90790-ABR94016)
CC for herbicidally active compounds, comprising aligning and comparing
CC nucleic acid or amino acid sequences from plant with nucleic acid or
CC amino acid sequences from non-plant organisms using suitable search
CC parameters, where plant sequences having an E-value greater by a factor
CC of 3 than the E-value of most similar non-plant sequences are selected.
CC The polypeptides or nucleic acids encoding them are useful for
CC identifying modulators. The identified modulators are useful as
CC herbicides
XX
XX Sequence 360 AA;
SQ
Query Match 8.7%; Score 86; DB 5; Length 360;
Best Local Similarity 26.5%; Pred. No. 1.8; Mismatches 20; Gaps 9;
Matches 49; Conservative 23; Indels 20; Gaps 9;
QY 7 SNLSNLTVDASLNGV--DKXLLSAEYKMLVQKADNEG1---EYVFEULLLYALAAR 60
DB 137 SEQPMFMSYIARLKGIVGDKKAMEIINNAFVWSAGPNDFLNYEIPSRRLPYPTSG 196
QY 61 TTSPVQVRDSDVITSNSRGERNVVVTGDDLKKVLDGAPLRFNKLRTGRTFTEAYV 120
DB 197 YDDFLTKLENFVRELYSIGVRNVLY--GGLPPM--GCLPI-HMTAKRNIFFRCLEHNN 251
QY 121 DECIAYKHLPLQJNAAAEIGIPEADSYLAADFLGTCPKLSEIQSRKMFASMYALK-TEG 179
DB 252 KDSVLYNEKLGWLLPQIEASLPG-SKFLYADVYN--PMMEMIQNPSSK-----YGFKETKR 303
QY 180 GVVNT 184
DB 304 GCCGT 308
RESULT 10
AAAG39165
ID AAAG39165 standard; protein; 366 AA.
XX
XX AAAG39165;
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 48419.
DE
XX Arabidopsis thaliana protein fragment SEQ ID NO: 48419.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridization assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX
XX Arabidopsis thaliana:
OS
XX
XX EP1033405-A2.
PN
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-00301439.
PF
XX
XX 25-FEB-1999; 99US-0121825P.
PR 25-FEB-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR

PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0125264P.
PR 28-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 22-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132487P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 18-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135153P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137528P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138547P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 18-JUN-1999; 99US-0139817P.
PR 21-JUN-1999; 99US-0139899P.
PR 22-JUN-1999; 99US-0140053P.
PR 23-JUN-1999; 99US-0140054P.
PR 23-JUN-1999; 99US-0140055P.
PR 24-JUN-1999; 99US-0140056P.
PR 24-JUN-1999; 99US-0140057P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141642P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142977P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.

PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144335P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144332P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145132P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 26-JUL-1999; 99US-0145919P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147319P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147433P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 13-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154033P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 26-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.

PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158233P.
PR 12-OCT-1999; 99US-0158368P.
PR 13-OCT-1999; 99US-0159233P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159328P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161924P.
PR 28-OCT-1999; 99US-0161939P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 8.7% Score 86; DB 3; Length 366;
Best Local Similarity 26.5%; Pred. No. 1.8; Matches 49; Conservative 23; Mismatches 93; Indels 20; Gaps 9;

QY 7 SNLSNLVITTDASSLNGV--DKKLISAIVEKMLVQKGFNEGI---EYVFGLLIYALNAR 60
DB 143 SEQPMNFKSYIARLKGIVGDKKAMEIINNAPVVSAGHNDFLINYYEIPSRLEYPFISG 202
QY 61 TTSFKVQADSDVITSNFGERNVYVTEGDLKKVLDGCAPLTRFNKJRTGRTTEAYV 120
DB 203 YODFTUKLENFVRRLYSIGVAVLV--GGLPPM--GCLPI-HMTAKRNIIFRPLEHHN 257
QY 121 DPCIAVKHKLPOLNNAASIGIPAEBSYLAADFLGTCPKLSLQGRKKVPSMAYALK-TEG 179
DB 258 KDSVLNEXLQVLLPQIEASLPG-SKFLYADVYN--PMMEMIQNPSK-----YGFKEKTKR 309
QY 180 GVVNT 184
DB 310 GCCGT 314

RESULT 11
AAG20733
ID AAG20733 standard; protein; 366 AA.
XX AAG20733;
AC
XX 17-OCT-2000 (first entry)
DT
XX Arabidopsis thaliana protein fragment SEQ ID NO: 23035.
DE
XX Arabidopsis thaliana protein fragment SEQ ID NO: 23035.
KM Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX Arabidopsis thaliana.
OS
XX EP1033405-A2.
PN
XX 06-SEP-2000.
PD
XX

Fri Apr 9 15:30:07 2004

us-09-613-486-15.rag

Page 12

PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P
PR 05-MAR-1999; 99US-0123180P
PR 09-MAR-1999; 99US-0123548P
PR 23-MAR-1999; 99US-0125788P
PR 25-MAR-1999; 99US-0126264P
PR 29-MAR-1999; 99US-0126785P
PR 01-APR-1999; 99US-0127462P
PR 06-APR-1999; 99US-0128234P
PR 08-APR-1999; 99US-0128744P
PR 16-APR-1999; 99US-0129845P
PR 19-APR-1999; 99US-0130077P
PR 21-APR-1999; 99US-0130449P
PR 23-APR-1999; 99US-0130510P
PR 28-APR-1999; 99US-0131449P
PR 30-APR-1999; 99US-0132048P
PR 30-APR-1999; 99US-0132407P
PR 04-MAY-1999; 99US-0132484P
PR 05-MAY-1999; 99US-0132485P
PR 06-MAY-1999; 99US-0132486P
PR 06-MAY-1999; 99US-0132487P
PR 07-MAY-1999; 99US-0132863P
PR 11-MAY-1999; 99US-0134256P
PR 14-MAY-1999; 99US-0134218P
PR 14-MAY-1999; 99US-0134219P
PR 14-MAY-1999; 99US-0134221P
PR 14-MAY-1999; 99US-0134370P
PR 18-MAY-1999; 99US-0134768P
PR 19-MAY-1999; 99US-0134941P
PR 20-MAY-1999; 99US-0135124P
PR 21-MAY-1999; 99US-0135253P
PR 24-MAY-1999; 99US-0135629P
PR 25-MAY-1999; 99US-0136021P
PR 27-MAY-1999; 99US-0136392P
PR 28-MAY-1999; 99US-0136782P
PR 01-JUN-1999; 99US-0137222P
PR 03-JUN-1999; 99US-0137528P
PR 04-JUN-1999; 99US-0137502P
PR 07-JUN-1999; 99US-0137724P
PR 08-JUN-1999; 99US-0138094P
PR 10-JUN-1999; 99US-0138540P
PR 10-JUN-1999; 99US-0138847P
PR 14-JUN-1999; 99US-0139119P
PR 16-JUN-1999; 99US-0139452P
PR 16-JUN-1999; 99US-0139455P
PR 17-JUN-1999; 99US-0139452P
PR 18-JUN-1999; 99US-0139454P
PR 18-JUN-1999; 99US-0139455P
PR 18-JUN-1999; 99US-0139456P
PR 18-JUN-1999; 99US-0139457P
PR 18-JUN-1999; 99US-0139458P
PR 18-JUN-1999; 99US-0139459P
PR 18-JUN-1999; 99US-0139460P
PR 18-JUN-1999; 99US-0139461P
PR 18-JUN-1999; 99US-0139462P
PR 18-JUN-1999; 99US-0139463P
PR 18-JUN-1999; 99US-0139750P
PR 18-JUN-1999; 99US-0139763P
PR 21-JUN-1999; 99US-0139817P
PR 22-JUN-1999; 99US-0139899P
PR 23-JUN-1999; 99US-0140353P
PR 23-JUN-1999; 99US-0140354P
PR 24-JUN-1999; 99US-0140695P
PR 28-JUN-1999; 99US-0140823P
PR 29-JUN-1999; 99US-0141287P
PR 30-JUN-1999; 99US-0141842P
PR 01-JUL-1999; 99US-0142154P
PR 01-JUL-1999; 99US-0142055P
PR 06-JUL-1999; 99US-0142390P
PR 08-JUL-1999; 99US-0142803P

PR 09-JUL-1999; 99US-0142920P
PR 12-JUL-1999; 99US-0142977P
PR 13-JUL-1999; 99US-0143542P
PR 14-JUL-1999; 99US-0143624P
PR 15-JUL-1999; 99US-0144005P
PR 16-JUL-1999; 99US-0144085P
PR 16-JUL-1999; 99US-0144086P
PR 19-JUL-1999; 99US-0144325P
PR 19-JUL-1999; 99US-0144331P
PR 19-JUL-1999; 99US-0144332P
PR 19-JUL-1999; 99US-0144333P
PR 19-JUL-1999; 99US-0144334P
PR 19-JUL-1999; 99US-0144335P
PR 19-JUL-1999; 99US-0144352P
PR 20-JUL-1999; 99US-0144632P
PR 20-JUL-1999; 99US-0144884P
PR 20-JUL-1999; 99US-0144814P
PR 21-JUL-1999; 99US-0145086P
PR 21-JUL-1999; 99US-0145088P
PR 22-JUL-1999; 99US-0145087P
PR 22-JUL-1999; 99US-0145088P
PR 22-JUL-1999; 99US-0145089P
PR 22-JUL-1999; 99US-0145192P
PR 22-JUL-1999; 99US-0145193P
PR 23-JUL-1999; 99US-0145145P
PR 23-JUL-1999; 99US-0145218P
PR 23-JUL-1999; 99US-0145224P
PR 26-JUL-1999; 99US-0145276P
PR 27-JUL-1999; 99US-0145913P
PR 27-JUL-1999; 99US-0145918P
PR 27-JUL-1999; 99US-0145919P
PR 27-JUL-1999; 99US-0145951P
PR 28-JUL-1999; 99US-0146386P
PR 02-AUG-1999; 99US-0146388P
PR 02-AUG-1999; 99US-0146389P
PR 02-AUG-1999; 99US-0146390P
PR 03-AUG-1999; 99US-0147038P
PR 04-AUG-1999; 99US-0147204P
PR 04-AUG-1999; 99US-0147302P
PR 05-AUG-1999; 99US-0147192P
PR 05-AUG-1999; 99US-0147260P
PR 06-AUG-1999; 99US-0147303P
PR 06-AUG-1999; 99US-0147416P
PR 09-AUG-1999; 99US-0147935P
PR 09-AUG-1999; 99US-0148117P
PR 10-AUG-1999; 99US-0148117P
PR 11-AUG-1999; 99US-0148341P
PR 12-AUG-1999; 99US-0148341P
PR 13-AUG-1999; 99US-0148565P
PR 13-AUG-1999; 99US-0148684P
PR 16-AUG-1999; 99US-0149168P
PR 17-AUG-1999; 99US-0149175P
PR 18-AUG-1999; 99US-0149426P
PR 20-AUG-1999; 99US-0149722P
PR 20-AUG-1999; 99US-0149723P
PR 20-AUG-1999; 99US-0149902P
PR 23-AUG-1999; 99US-0149902P
PR 23-AUG-1999; 99US-0149930P
PR 25-AUG-1999; 99US-0150566P
PR 26-AUG-1999; 99US-0150884P
PR 27-AUG-1999; 99US-0151065P
PR 27-AUG-1999; 99US-0151066P
PR 27-AUG-1999; 99US-0151080P
PR 30-AUG-1999; 99US-0151303P
PR 31-AUG-1999; 99US-0151430P
PR 01-SEP-1999; 99US-0151930P
PR 07-SEP-1999; 99US-0152363P
PR 10-SEP-1999; 99US-0153770P
PR 13-SEP-1999; 99US-0153758P
PR 15-SEP-1999; 99US-0154018P
PR 16-SEP-1999; 99US-0154039P
PR 20-SEP-1999; 99US-0154779P
PR 22-SEP-1999; 99US-0155139P
PR 23-SEP-1999; 99US-0155486P
PR 24-SEP-1999; 99US-0155659P

PR 28-SEP-1999; 99US-0156458P.
 PR 29-SEP-1999; 99US-0156596P.
 PR 04-OCT-1999; 99US-0157117P.
 PR 05-OCT-1999; 99US-0157753P.
 PR 06-OCT-1999; 99US-0157865P.
 PR 07-OCT-1999; 99US-0158029P.
 PR 08-OCT-1999; 99US-0158322P.
 PR 12-OCT-1999; 99US-0158369P.
 PR 13-OCT-1999; 99US-0159293P.
 PR 13-OCT-1999; 99US-0159294P.
 PR 13-OCT-1999; 99US-0159295P.
 PR 14-OCT-1999; 99US-0159329P.
 PR 14-OCT-1999; 99US-0159330P.
 PR 14-OCT-1999; 99US-0159331P.
 PR 14-OCT-1999; 99US-0159637P.
 PR 14-OCT-1999; 99US-0159638P.
 PR 18-OCT-1999; 99US-0159584P.
 PR 21-OCT-1999; 99US-0160741P.
 PR 21-OCT-1999; 99US-0160767P.
 PR 21-OCT-1999; 99US-0160768P.
 PR 21-OCT-1999; 99US-0160770P.
 PR 21-OCT-1999; 99US-0160814P.
 PR 21-OCT-1999; 99US-0160815P.
 PR 22-OCT-1999; 99US-0160980P.
 PR 22-OCT-1999; 99US-0160981P.
 PR 22-OCT-1999; 99US-0160989P.
 PR 22-OCT-1999; 99US-0161404P.
 PR 25-OCT-1999; 99US-0161405P.
 PR 25-OCT-1999; 99US-0161406P.
 PR 26-OCT-1999; 99US-0161359P.
 PR 26-OCT-1999; 99US-0161360P.
 PR 26-OCT-1999; 99US-0161361P.
 PR 28-OCT-1999; 99US-0161920P.
 PR 28-OCT-1999; 99US-0161922P.
 PR 28-OCT-1999; 99US-0161993P.
 PR 29-OCT-1999; 99US-0162142P.

Query March 8.7%; Score 86; DB 3; Length 366;
 Best Local Similarity 26.5%; Pred. No. 1.8;
 Matches 49; Conservative 23; Mismatches 93; Indels 20; Gaps 9;

QY 7 SNISNLTVDASSINGV--DKKLISAEEVKMLVQKAPNNGI-----EVPFGLLYLAAR 60
 DB 143 SEQPNMKSYIARLKGIVGDKKAMEIINNAFVVASAGPNDPILNYEIPSRRLYEYFISG 202
 QY 61 TTSPKVGRAADSDVYFNSFGERNVVTGDLKKVLDGCAPLTRFTNKLRTFGRTFEAVY 120
 DB 203 YQDFILKRLNFPRELKSLGVNVLV--GGLPRV--GCLPI-IMTKFRNIFRPECLEHNN 257
 QY 121 DFCIAYGKLPQLNAAALGIPADSYLADFLGTCKKSELQOSRMFMSTALK-TBG 179
 DB 258 KDSVLYNEKIQNLIPQIEASLPG-SKFLYADVNN--FMEMIQNPSK-----YGFRETKR 309
 QY 180 GVYNT 184
 DB 310 GCCGT 314

RESULT 12
 AAB96072
 ID AAB96072 standard; protein: 447 AA.

XX AC AAB96072;
 XX DT 29-OCT-2001 (first entry)
 XX DE Putative P. abyssi beta-lactamase.
 XX KM Hyperthermophilic archaeon; hyperthermophilic protein.
 XX OS Pyrococcus abyssi.
 XX PN FR2792651-A1.

XX PD 27-OCT-2000.
 XX PF 21-APR-1999; 99FR-00005034.
 XX PR 21-APR-1999; 99FR-00005034.
 XX PA (CNRS) CNRS CENT NAT RECH SCI.
 XX PI (IFRE)-IFREMER INST FR RECH EXPL MER.
 XX PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O,
 XX PI Querellou J, Weissenbach J, Saurin W, Helling R;
 XX DR WPI; 2001-126236/14.
 XX PT New nucleotide sequences isolated from Pyrococcus abyssi encode proteins
 XX PS useful in industry.
 XX CC Claim 7; Page 693-694; 1657bp; French.
 XX CC The present invention relates to the genomic sequence of Pyrococcus
 XX CC abyssi (see AB86431 and AAH41223-7) and P. abyssi proteins. P. abyssi is
 XX CC a hyperthermophilic archaeon, which is isolated from deep-sea
 XX CC hydrothermal vents. The present sequence is one such P. abyssi protein.
 XX CC The proteins of the present invention have various potential industrial
 XX CC uses, since the proteins are stable at very high temperatures, some up to
 XX CC 110 degrees centigrade. Note: This patent is in the same patent family as
 XX CC WO200005062, which contains additional sequences as shown in AAB99132-
 XX CC AAB99143, AAH75903-AAH75920 and AAG66436
 XX SQ Sequence 447 AA;

Query March 8.7%; Score 86; DB 4; Length 447;
 Best Local Similarity 26.0%; Pred. No. 2.4;
 Matches 40; Conservative 15; Mismatches 41; Indels 58; Gaps 8;

QY 26 KILSAEVRKMLVQKAPNNGIEVVPFGLLYLAARTSPKVRADSDVIFNSFGERNV- 84
 DB 5 KLESPFVKMERK-VPGISISII-----KGDVYAKGQYRIVE 44
 QY 85 -----VTEGDKKVLGCAPLTRFTN-XLRTPGRTFT 116
 DB 45 ARLEPSPETIYGISITSPFLALIMKLVBEGLS--LDD--PYEKFNKLRPFGEPRVT 100
 QY 117 EAVYDFCIAYGKLPQLNAAAL--GIPADSYL 148
 DB 101 ---VAHLTLHSSGIFSLGVAEAFIDGVVGDNDWL 131

RESULT 13
 AAH17788
 ID AAH17788 standard; protein: 599 AA.

XX AC AAH17788;
 XX DT 17-OCT-2003 (revised)
 XX DT 19-AUG-1997 (first entry)
 XX DE Phage abortive infection protein AbiE.
 XX KM AbiE; abortive phage infection protein; phage resistance; pSRQ800;
 XX KM lactic acid bacterium.
 XX OS Lactococcus lactis subsp. lactis; isolate W1.
 XX PN WO9720917-A2.
 XX DT 12-JUN-1997.
 XX PF 20-NOV-1996; 96WO-IB001385.
 XX PR 01-DEC-1995; 95US-00565907.

PA (UNIL) QUEST INT BV.
 PI Moineau S, Holler BJ, Vandenberg PA, Vedamuthu ER, Kondo JK;
 XX N-PSDB; AAT68648.
 DR WPI; 1997-319765/29.
 XX
 PT Isolated DNA encoding the Abie protein of Lactococcus - for protecting
 PT strains used in production of fermented dairy products.
 PS Claim 48; Page 30-32; 49pp; English.
 XX
 CC Abie (AAM17788) is a phage abortive infection protein that increases
 CC resistance to phage. It is the expression product of an open reading
 CC frame found in the 4.5 kb ECORI fragment (AAT68648) of Lactococcus lactis
 CC (L.L.) subsp. lactis W1 plasmid pSRQ800. Abie acts at, or before, phage
 CC replication and has no homology with known Abi proteins. It protects
 CC against phages of 936 and P335 types, also against C2 when present on a
 CC high copy number plasmid. Abie can be used to impart phage resistance to
 CC bacteria, esp. L.L. that do not have natural resistance. The recombinant
 CC bacteria can be used in the prodn. of fermented dairy products. (Updated
 CC on 17-OCT-2003 to standardise OS field)
 XX
 SQ Sequence 599 AA;
 Query Match 8.6%; Score 85; DB 2; Length 599;
 Best Local Similarity 24.3%; Pred. No. 4.7;
 Matches 36; Conservative 26; Mismatches 46; Indels 40; Gaps 6;
 QY 2 ELMSDSNL-----SLVITD-----ASSINGVDK--KLSA 30
 DB 261 EFLERFNLICENBMLINDNKTVDNPPYDKSKSDIFSFENITSTNDKMKIEISN 320
 QY 31 EWEKMLVQKAPNEG-IEVFGLLLYALARTSPKQPADSDVIFSNSEGERVAVTEG 89
 DB 321 FIDYCVNEEHLKNGAKIKCIPPI-----TNTLKQKQVDTKNI-DNIFSKRMVTFNEN 372
 QY 90 DLKRYLDGCAPLRFPTNKLRTFGRTFTE 117
 DB 373 VFEKILDLKDSKRLINKFTFFENINE 400
 RESULT 14
 AAY97690
 ID AAY97690 standard; protein; 269 AA.
 XX
 AC AAY97690;
 XX
 DT 06-AUG-2003 (revised)
 DT 08-MAY-2001 (first entry)
 XX
 DE GRAY-5 coat protein.
 XX
 KW GRAY-5; grapevine leafroll virus; GRAY infection; GRAY coat protein;
 KM GRAY HSP70 homologue protein; viral gene mapping;
 KM plant disease resistance.
 XX
 OS Grapevine leafroll virus.
 XX
 FN WO200105957-A2.
 XX
 PD 25-JAN-2001.
 XX
 PF 19-JUL-2000; 2000WO-US019708.
 XX
 PR 19-JUL-1999; 99US-0144453P.
 XX
 PA (AGRI-) AGRITOPIC INC.
 XX
 PI Good XC, Monis J;
 XX
 DR WPI; 2001-147339/15.
 DR N-PSDB; AAA91259.

XX
 PT Novel grapevine leafroll virus polynucleotide useful as diagnostic and
 XX probe, for viral gene mapping and for induced plant disease resistance.
 XX
 PS Claim 1; Fig 1; 60pp; English.
 XX
 CC This sequence represents a grapevine leafroll virus (GRAY-5) protein
 CC sequence of the invention. The DNA sequence can be used in an expression
 CC construct. The construct is useful for providing resistance to GRAY
 CC infection in a recombinant plant cell by transforming the plant cell with
 CC it, where transcription of the polynucleotide sequence interferes with a
 CC normal viral function such as movement, encapsidation or replication of
 CC viral RNA. The polynucleotide sequence is expressed as an antisense
 CC sequence and encodes a GRAY coat protein, preferably a defective GRAY
 CC coat protein or a GRAY HSP70 homologue protein. The GRAY-5 DNA is
 CC useful for the synthesis of GRAY, as diagnostic and probes, for viral
 CC gene mapping and for induced plant disease resistance. It is also useful
 CC to detect and quantitate expression of GRAY in plant tissue prior to use
 CC in vegetative propagation, by detecting the presence of GRAY RNA.
 CC (Updated on 06-AUG-2003 to correct OS field.)
 XX
 SQ Sequence 269 AA;
 Query Match 8.5%; Score 84.5; DB 4; Length 269;
 Best Local Similarity 24.4%; Pred. No. 1.7;
 Matches 32; Conservative 21; Mismatches 57; Indels 21; Gaps 4;
 QY 61 TTSPKQVQADSDVIFSNSEGERVAVTEGDKVLDGCAPLRFPTNKLRTFGRTFTEAVY 120
 DB 127 STSPKQVQADSDVIFSNSEGERVAVTEGDKVLDGCAPLRFPTNKLRTFGRTFTEAVY 185
 QY 121 DFCIAVYKHLPLNAAAEISGIPAE-----DSYLDPLGTCPKSEL 162
 DB 186 QGISGKLEV-NTKICASHGVPPNYFPSPDLHVDAFLFGYDASLAEL-GRKVAIVNKP 243
 QY 163 QOSRCKMFMAMY 173
 DB 244 SNSNRATNLY 254
 RESULT 15
 ABU26015
 ID ABU26015 standard; protein; 422 AA.
 XX
 AC ABU26015;
 XX
 DT 19-JUN-2003 (first entry)
 DT XX
 DE Protein encoded by prokaryotic essential gene #11542.
 XX
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX
 OS Corynebacterium diphtheriae.
 XX
 PN WO200277183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362599P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind UT,
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 DR WPI; 2003-029926/02.
 DR N-PSDB; ACA29885.

XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.

PS Claim 25; SEQ ID NO 53939; 1766bp; English.

XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

CC
XX
SQ Sequence 422 AA;

Query Match 8.4%; Score 83.5; DB 6; Length 422;

Best Local Similarity 24.6%; Pred. No. 4.2;

Matches 46; Conservative 15; Mismatches 67; Indels 59; Gaps 7;

QY	53	LVYALARTTSPKVQRA	SVIFSN	SFGERNVV	TBEGDL	KKVLDGCA	PLTRFTNK	RTF-	111
DB	106	LIRPFAETTDSTQTA	HDTRVFC	NSGTEAN	-----	EA	FKLARMTG	KNRILA	153
QY	112	-----	GRTF	-----	TEAYVDF	CIAYKH	KL	POLNAA	-----
DB	154	AQGHFGRITGALAM	TGQPKOMP	FYPLPAGV	EFTYFGDID	YLRKLVQ	INSTVAA	IIIE	213
QY	137	---AELG-IP	AEDSYLAD	FLGTCPK	-----	LSEIQ	SRKMPAS	MYALKT	BEGVVNTPY
DB	214	PIQETGVIPAPD	FLISA-V	RGLCN	EPDILMI	VDVQ	TGIGRTG	DFP	QHETGVIPDVY
QY	187	SNLRQLG	193						
DB	273	TMARGLG	279						

Search completed: April 5, 2004, 17:52:41
Job time : 62 secs

C/Accession: S28715
 R/Agarose, A.A.; Boyko, V.P.; Karasev, A.V.; Lunina, N.A.; Koonin, E.V.; Dolja, V.V.
 J/Gen. Virol. 72, 15-23, 1991
 A/Title: Nucleotide sequence of the 3'-terminal half of beet yellows closterovirus RNA
 A/Reference number: S28710; MUID:9116305; PMID:1990061
 A/Accession: S28715
 A/Molecule type: DNA
 A/Residues: 1-204 <AGR>
 A/Cross-references: EMBL:X53462; NID:G58878; PID:CA37554.1; PID:G58883
 C/Superfamily: SBYV probable coat protein

Query Match 28.9%; Score 286; DB 2; Length 204;
 Best Local Similarity 37.6%; Pred. No. 3e-19;
 Matches 71; Conservative 29; Mismatches 87; Indels 2; Gaps 2;

QY 7 SNLSNLVITPDASLNGVKKLLSAVEKMLVOKAPNGIEVFGLLYLAARATSPKV 66
 DB 11 ATFEVSLADQTCIHGEDCDKIKRFEBCLKIGVPEBNLIALGLCLYSCATTGTSNKV 70
 QY 67 QRAQSDVIFSNF-GERNVVVTEGDLKKVLDGCAPLRTFNKLRTPGRTTEAVYDCIA 125
 DB 71 NVQPTSTFKASFGGKELVLTGELNSFLGSKLSEKPKLRFCRPFCKDYISRKE 130
 QY 126 YKHKLPOLMAAELGIPEDSYLADFLGTCPKLSLQOSKRMASWALKTGGVNTP 185
 DB 131 YRGLPPIARNRHGLPAEDHYLADFLSTSTELTDLQOSRLARENATHTFS-SESP 189
 QY 186 VSNLRQLGR 194
 DB 190 VTSLKQLGR 198

RESULT 3
 D49804
 capsid protein p25 - citrus tristeza closterovirus CTV

C/Species: citrus tristeza closterovirus CTV
 C/Date: 11-Oct-1994 #sequence_revision 25-Apr-1997 #text_change 11-Jan-2000
 C/Accession: D49804
 R/Papou, H.R.; Karasev, A.V.; Anderson, E.J.; Papou, S.S.; Hilt, M.E.; Febres, V.J.; Eck
 S.M.; Dawson, W.O.; Lee, R.F.; Niblett, C.L.
 A/Title: Nucleotide sequence and organization of eight 3' open reading frames of the cit
 A/Reference number: A49804; MUID:94160579; PMID:8116253
 A/Accession: D49804
 A/Status: preliminary
 A/Molecule type: genomic RNA
 A/Residues: 1-223 <PAP>
 A/Note: sequence extracted from NCBI backbone (NCBIN:144092, NCBI:P:144102)
 A/Note: severe quick decline isolate T36
 C/Superfamily: SBYV probable coat protein

Query Match 12.6%; Score 125; DB 2; Length 223;
 Best Local Similarity 25.9%; Pred. No. 0.0036;
 Matches 56; Conservative 33; Mismatches 63; Indels 64; Gaps 10;

QY 18 SLSN-GVKKLLSA-EVERKMLVOKA-----PNEGIEVFGLLYLA 58
 DB 29 SSVNHIPTLTITMDVQLSTQNAALNRDLFLTKGKHPLPDKDCHFIAMLYELA 88
 QY 59 ARTTS-----PKYQRAD--SDVIF-SNSFGERNVVTEBDLKKVLDGC 98
 DB 89 VSSSLQSDDDATGITYREGVEVLDLMDTDVFNKSGIGNR-----132
 QY 99 APLTFNKLRTFGRTTEAVYDPCIAVYKLPOLMAAELGIPEDSYLADFLGTCPK 158
 DB 133 -----TNALRVGSGTNDALYAF-C-RQNRNLSYGRPLDAGIPAGHYLCADEL-TRAG 184
 QY 159 LSELQOSKRMASWALKTGGVNTPVSNLQOLGR 194
 DB 185 LTDLCAVYIQKEQLK-KRGADVVVTVNQLGR 219

RESULT 4

S27899
 hypothetical protein 2 - sugar beet yellows virus
 C/Species: sugar beet yellows virus, SBYV
 C/Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 10-Dec-1999
 C/Accession: S27899
 R/Brumsted, J.; Moseley, J.; Hull, R.
 A/Title: Nucleotide sequence of cDNA encoding the coat protein of beet yellows viru
 A/Reference number: S27898
 A/Accession: S27899
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-216 <BRU>
 A/Cross-references: EMBL:M59452; NID:G323237; PID:AAA72954.1; PID:G323239
 C/Superfamily: SBYV probable coat protein

Query Match 10.5%; Score 104; DB 2; Length 216;
 Best Local Similarity 28.5%; Pred. No. 0.032;
 Matches 41; Conservative 19; Mismatches 62; Indels 22; Gaps 6;

QY 61 TTSPPVQRAQSDVIFSNFSGERNVVVTEGDLKKVLDGCAPLRTFNKLRTPGRTTEAVY 120
 DB 83 TTSKTVYVGA---YEYTIQKKFLVKDAMVPPLIKECKKKNKPNVPTFCATFEDAY- 138
 QY 121 DPCIAVYKLPOLN--AAAEIGIPEDSYLADFL--GTC---PKLSLQOSKRMKA 170
 DB 139 --TVIARSLPKLNLNRTIGKRGIPSGYEFGLADFLTATVCLNDHEKAIYLAASRAID 195
 QY 171 SMYALKTGGVNTPVSNLQOLGR 194
 DB 196 RAYSSSVDGKIV-----SLFDLGR 214

RESULT 5

S34205
 coat protein homolog - sugar beet yellows virus
 C/Species: sugar beet yellows virus, SBYV
 C/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 10-Dec-1999
 C/Accession: S34205
 R/Agarose, A.A.; Koonin, E.V.; Boyko, V.P.; Maize, E.; Lunina, N.A.; Atabekov, J.G.
 A/Title: Nucleotide sequence and organization of the Ukrainian isolate of Beet yellows viru
 A/Reference number: S34202
 A/Accession: S34205
 A/Status: preliminary
 A/Molecule type: genomic RNA
 A/Residues: 1-216 <AGR>
 A/Cross-references: EMBL:X73475; NID:G313689; PID:CA451860.1; PID:G313693
 C/Superfamily: SBYV probable coat protein

Query Match 10.1%; Score 100.5; DB 2; Length 216;
 Best Local Similarity 25.0%; Pred. No. 0.067;
 Matches 50; Conservative 27; Mismatches 94; Indels 29; Gaps 7;

QY 5 SDSLNSNLVITPDASSLNGVKKLLSAVEKMLVOKAPNGIEVFGLLYLAARATSP 64
 DB 34 SEVNPKNLKRKETDELGVIRERFSEIV-----ITDEDVFKHAFALIPAAITIST 86
 QY 65 KVQRAQSDVIFSNFSGERNVVVTEGDLKKVLDGCAPLRTFNKLRTPGRTTEAVYDCI 124
 DB 87 KVVYVGA---YEYTIQKKFLVKDAMVPPLIKECKKKNKPNVPTFCATFEDAY----I 139
 QY 125 AYVHKLPOL--NAAAEIGIPEDSYLADFL--GTC---PKLSLQOSKRMASWKA 174
 DB 140 VIARSLPKLNLNRTIGKRGIPSGYEFGLADFLTATVCLNDHEKAIYLAASRAIDRAVS 199
 QY 175 LKTEGGVNTPVSNLQOLGR 194
 DB 200 SSVDGKIV-----SLFDLGR 214

RESULT 6
 C49804

capsid protein homolog p27 - citrus tristeza closterovirus CTV

C:Species: citrus tristeza closterovirus CTV
C>Date: 11-Oct-1994 #sequence_revision 25-Apr-1997 #text_change 11-Jan-2000

C/Accession: C49804

R/Pappu, H.R.; Karasev, A.V.; Anderson, E.J.; Pappu, S.S.; Hilt, M.E.; Febres, V.J.; Eck S.M.; Dawson, W.O.; Lee, R.F.; Mblett, C.L.

Virology 199, 35-46, 1994

A>Title: Nucleotide sequence and organization of eight 3' open reading frames of the cit

A/Reference number: A49804; MUID:94160579; PMID:8116253

A/Accession: C49804

A/Status: preliminary

A/Molecule type: genomic RNA

A/Residues: 1-240 <PDP>

A/Cross-references: GB:U16304; GB:U02547; GB:L20760; NID:G806738; PIDN:AA059629.1; PID:9

A/Note: sequence extracted from NCBI backbone (NCBIN:144092, NCBI:P144099)

A/Note: severe quick decline isolate T36

C/Superfamily: SBV probable coat protein

Query Match 10.1%; Score 100.5; DB 2; Length 240;

Best Local Similarity 26.4%; Pred. No. 0.077;

Matches 57; Conservative 28; Mismatches 70; Indels 61; Gaps 13;

2 EIMSDSLNVLIT-DASLNGVDKLLSAEVEKMLVOKGAPNKGIEVVGGLLYALAA 60

61 EKFTGEHLKRYVMVMDTFLPENTKTEDLLVLTMTQK-----RLYIT--- 104

62 TTSPKRVADSDVIFSNFGRNVVTEGDLK-KVLDGCA--PLTRFT----NKLRTF 111

105 STSKTKTFRDKGCI-----SVQGGRLRYLLDKVPEPFIKSTRETREPNAIRK 154

112 GRTTEAVDPCIA-----YKHLPLQMAAELGIPAEBSYLAADL-CTCKLER--- 161

155 ACTEEELHL--CMARLPDLYENK---RTTRAGTGLKGLYSADPLSGISSEHER 207

162 ---LQGRKMFASVYALKTEGGVNTFVSNLRQGR 194

208 GILIRASMSMA-----RQGYEATLNLNRDGLK 238

RESULT 7

528714

hypothetical protein 5 - sugar beet yellows virus

C/Species: sugar beet yellows virus, SBV

C/Date: 07-May-1993 #sequence_revision 07-May-1993 #text_change 10-Dec-1999

C/Accession: S28714

R/Aganovsky, A.A.; Boyko, V.P.; Karasev, A.V.; Lunina, N.A.; Koonin, E.V.; Dolja, V.V.

J. Gen. Virol. 72, 15-23, 1991

A>Title: Nucleotide sequence of the 3'-terminal half of beet yellows closterovirus RNA

A/Reference number: S28710; MUID:9116305; PMID:1990061

A/Accession: S28714

A/Molecule type: DNA

A/Residues: 1-216 <AGR>

A/Cross-references: EMBL:X53462; NID:958678; PIDN:CAA31553.1; PID:958682

C/Superfamily: SBV probable coat protein

Query Match 10.0%; Score 99.5; DB 2; Length 216;

Best Local Similarity 25.0%; Pred. No. 0.083;

Matches 50; Conservative 27; Mismatches 94; Indels 29; Gaps 7;

5 SDSNLSNVLITDASLNGVDKLLSAEVEKMLVOKGAPNKGIEVVGGLLYALAA 64

34 SEVPNNKLNKRETDGLGVIRERKSELV-----ITDEDFVHLAFLIRANITTSV 86

65 KYQADSDVIFSNFGRNVVTEGDLK-KVLDGCAPLTRFTKRTGRTFTEAVVDFCI 124

87 KVNAYGA---YEVITGKKFLVKNAMVPEPLKECKKFNKNPVRTCAFPEDAY---- 139

125 AYKHLPLQMAAELGIPAEBSYLAADL--GTC---PLSLQGRKMFASMYA 174

140 VIASLPLFLNRTIGKRGIPSGEFLGADPLTATSVCLNDHEKAVIQLASRAIDRAVS 199

175 LKTEGGVNTFVSNLRQGR 194

DB 200 SSVDKIV-----SLFDLGR 214

RESULT 8

A33988

adenylate cyclase (EC 4.6.1.1) - fission yeast (Schizosaccharomyces pombe)

C/Species: Schizosaccharomyces pombe

C/Date: 23-Mar-1990 #sequence_revision 23-Mar-1990 #text_change 31-Mar-2000

C/Accession: A33988; A33539; T39809

R/Young, D.; Riggs, M.; Field, T.; Vojtek, A.; Broek, D.; Wiegler, M.

Proc. Natl. Acad. Sci. U.S.A. 86, 7989-7993, 1989

A>Title: The adenylate cyclase gene from Schizosaccharomyces pombe.

A/Reference number: A33988; MUID:90046723; PMID:2682634

A/Accession: A33539

A/Molecule type: DNA

A/Residues: 1-1692 <YOU>

A/Cross-references: GB:M26699; NID:G17338; PIDN:AA35284.1; PID:G17339

R/Yamawaki-Kataoka, Y.; Tamaoki, T.; Choe, H.R.; Tanaka, H.; Kataoka, T.

Proc. Natl. Acad. Sci. U.S.A. 86, 5693-5697, 1989

A>Title: Adenylate cyclases in yeast: a comparison of the genes from Schizosaccharomyces

A/Reference number: A33539; MUID:89345533; PMID:2669944

A/Accession: A33539

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-1692 <YAM>

A/Cross-references: GB:M24942; NID:G173378; PIDN:AA435301.1; PID:G173379

A/Note: the authors translated the codon TGC for residue 626 as Ser, and GCC for residue

R/Lyne, M.; Rajandream, M.A.; Bartell, B.G.; Lucas, M.; Galliardun, C.

submitted to the EMBL Data Library, June 1998

A/Reference number: Z21881

A/Accession: T39809

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1692 <LYN>

A/Cross-references: EMBL:AL023859; PIDN:CAA19571.1; GSPDB:GN00067; SPDB:SPBC19C7.03

A/Experimental source: strain 972h; cosmid c19C7

C/Genetics: SPBC19C7.03

A/Map position: 2

C/Superfamily: yeast adenylate cyclase; leucine-rich alpha-2-glycoprotein repeat homolog

C/Keywords: phosphorus-oxygen lyase

F/1328-1413/Domain: yeast adenylate cyclase catalytic domain homology <YACC>

Query Match 9.3%; Score 92; DB 2; Length 1692;

Best Local Similarity 24.7%; Pred. No. 5.9;

Matches 42; Conservative 26; Mismatches 66; Indels 34; Gaps 7;

27 LLSAEVEKMLVOKGAPNKGIEVVGGLLYALAAFTSPKVRADSDVIFSNFGRNVV 86

339 LIQFTERTIL-----PHEQPCIIFFRLISLFCCKYTS-----DEINIEDNYSVARLVF 388

87 TEGDLKVVLDGCAPLTRFNKLR-----FGRTFAVDFCIAYGKLPOLNAAELGIP 142

389 TMDI-----GADVLRKRSKKTITANLDSNLEVIPIKTIYPYHLLISNLSNLSLD 443

143 AEDSYLAADFLCTCKLSLQGRKMFASVYALKTEGGVNTFVSNLRQGR 192

444 -----LPIDFMERCVALKRLDISNN-----LRSPRG--KFTALRQL 478

RESULT 9

AD3246

P-450 monoxygenase vtrH [imported] - Agrobacterium tumefaciens (strain C58, Dupont) pl.

C/Species: Agrobacterium tumefaciens

C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002

C/Accession: AD3246

R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Moks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.

erage, G.; Giller, W.; Grant, C.; Guenther, D.; Kuyavlin, T.; Levy, R.; Li, M.; McCall, L.

Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, I.

ster, E.W.

A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; MWID:21608550; PMID:11743193

A:Accession: AD3246
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-419 <KIR>
A:Cross-references: GB:AE008690; PIDN:AAL46386.1; PID:g17744179; GSPDB:GN00189
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: vlrH

A:Genome: plasmid
C:Superfamily: Agrobacterium plasmid cytochrome P450 pinF; cytochrome P450 homology
C:Keywords: heme; iron; metalloprotein
F:366/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 9.1%; Score 90; DB 2; Length 419;

Best Local Similarity 25.1%; Pred. No. 1.5; Mismatches 71; Indels 58; Gaps 11;

Matches 52; Conservative 26; Mismatches 71; Indels 58; Gaps 11;

Db 20 LMGVKKLLSAEYKMLVOKAPNBEIE--VFGLLYLAARTSPKQVQADSDVIFSN 77

Qy 67 LMGVKKLLSAEYKMLVOKAPNBEIE--VFGLLYLAARTSPKQVQADSDVIFSN 77

Db 78 SFGRRNVVTEGDLKKVLDGAP-LTRFTNKLRTFGFTFEAVDPICIAKHLPOLNA 136

Qy 118 TFAFR-----MIDALRPEITKLTHTLMDVPRVDD--PDEAKMTASKLPALITL 164

Db 137 AELGIPAD-----SYLADFLGTCPKLSHQ-----QSRKM-- 168

Qy 165 SVLSPFGDAPFFTRIVYNSRCLSPMGEDDPFEISASVELQDYVRAVADRSRRISD 224

Db 169 -FASNY--AKTGGVANTPVSNLROJ 192

Qy 225 DFLSCYLKAVREEGTL--SPTEIMOL 249

Db 225 DFLSCYLKAVREEGTL--SPTEIMOL 249

RESULT 10

phosphoenolpyruvate carboxykinase pckA - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 27-Oct-2003

C:Accession: F69673

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berber, C.; Bron, S.; Brouillette, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chik, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Extington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, C.R.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gallier, J.; Harwood, C.R.; Hensaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Huilo, M.F.

Koester, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Laubert, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Portetella, Rieger, M.; Rivoira, C.; Rooha, E.; Roche, B.; Rose, M.; Sadie, Y.; Sato, T.; Scanlon, A.; Schleich, S.; Schroeder, R.; Scoffone, F.; Scoffone, F.; Tognoni, J.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K.

A:Authors: Yoshikawa, H.F.; Zumestein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A:Reference number: A69580; MWID:98044033; PMID:9384377

A:Accession: F69673

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-557 <KUN>

A:Cross-references: GB:I29119; GB:AL009126; NID:G2635411; PIDN:CAH15034.1; PID:G2635540

A:Experimental source: strain 168

C:Genetics: pckA

C:Superfamily: phosphoenolpyruvate carboxykinase [ATP]

Db 234 LSGTKTLLSADNRKLLGDERGWSDTGVNIEG-GCYAKCIHLSBEK---EPQIFNA 288

Qy 77 NSFGE--RVNVTEGDLKKVLDGAP-LTRFTNKLRTFGFTFEAVDPICIAKHLPOL 133

Db 289 IRFGSVLENVYVDD-----TRANVDDSYTENTTA-----AYPLHMINNI 330

Qy 134 NAAELGIPADSYLADFLGTCPKLSHQ-----QSRKM--FASMYALK--TEGGVNV 183

Db 331 VTPSMAGHPSAIVPLTADAFGLVPLISKLTKEQVYHFLSGYTSKLAGTERGVT 385

RESULT 11

probable GDSL-motif lipase/hydrolase [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001

C:Accession: B84722

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MWID:20083487; PMID:10617197

A:Accession: B84722

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-328 <STO>

A:Cross-references: GB:AE002093; NID:G4582450; PIDN:AAD24834.1; GSPDB:GN00139

C:Genetics:

A:Gene: Atg31550

A:Map position: 2

C:Superfamily: myosinase-associated protein MyAP

Query Match 8.7%; Score 86.5; DB 2; Length 328;

Best Local Similarity 25.8%; Pred. No. 2.3; Mismatches 86; Indels 37; Gaps 10;

Matches 49; Conservative 18; Mismatches 86; Indels 37; Gaps 10;

Db 14 ITDASSLNGV-----DKKLSAEYKMLVOKAPNBEIEVFGLLYLAARTSP 64

Qy 105 LSDQDILITGCPASAGIYQDKKAMEIINNAFVVSAGBND-----FINNYDIFSRLEY 159

Db 65 KVGQADSDVIFSN-----SFGRRNVVTEGDLKKVLDGAP-LTRFTNKLRTFGFTF 115

Qy 160 PFIQGVDFILRLNFVRELISLVGRVNV--GGLPVW--GCLPI-HMTAKFRIRFFC 214

Db 116 TEAVVDFCIAYGKLPOLNAAELGIPADSYLADFLGTCPKLSHQ-----QSRKM--FASMYALK--TEGGVNV 175

Qy 215 LEHNNDSVLYNEKQKLPQIEASLPG-SKFLVADVNV--PMWEMIONPSK-----YGF 266

Db 176 K-TEGGVNV 184

Qy 267 KETKRCGCGT 276

Db 267 KETKRCGCGT 276

RESULT 12

probable GDSL-motif lipase/hydrolase [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001

C:Accession: B84722

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;

A:Map position: 2
C:Superfamily: myrosinase-associated protein MYAP

Query Match 8.7%; Score 86; DB 2; Length 360;
Best Local Similarity 26.5%; Pred. No. 2.9;
Matches 49; Conservative 23; Mismatches 93; Indels 20; Gaps 9;

QY 7 SNLSNLTVDASSLNGV--DKLLSAVEKMLVOKGAPNGI-----EVVGLLLYALAA 60
DB 137 SEQPMFKSYIARLKGIVGDKAMEIINNAFVVVAGPNDFILNYEIPSRRLYEPTISG 196
QY 61 TTSPKVRADSDVFNSSFGERNVVTGDLKVLDCAPLTFPTNKLRTFGRTFETAVY 120
DB 197 YQDPIKLRLNPFVLEISLGVNVLV--GGLPRM--GCLPI--HMTAKFRNIFPFCLEHN 251
QY 121 DFCIAYKHLFOLNAAALGIPADSYLADFLGTCRKLSELOGRKMFASVALK--TEG 179
DB 252 KDSVLYNKKLQNLPLQIEASLPD--SKFLYADVN--FMEMIQVPSK-----YGFETKR 303
QY 180 GVAVT 184
DB 304 GCCGT 308

RESULT 13

hypothetical protein T12A2.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T14352
R:Latreille, P.
submitted to the EMBL Data Library, March 1996
A:Description: The sequence of C. elegans cosmid T12A2.
A:Reference number: 221511
A:Accession: T14352
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-445 <LAT>
A:Cross-references: EMBL:U13019; PIDN:AAC24449.1; GSPDB:GN00021; CESP:T12A2.1
A:Experimental source: strain Bristol N2; clone T12A2
C:Genetics:
A:Gene: CESP:T12A2.1
A:Map position: 3
A:Introns: 33/3; 117/3; 146/3; 202/3; 250/2; 297/3; 371/1; 407/2

Query Match 8.7%; Score 86; DB 2; Length 445;
Best Local Similarity 24.7%; Pred. No. 3.9;
Matches 54; Conservative 41; Mismatches 80; Indels 44; Gaps 12;

QY 1 MEIISDSNLNLTVDAS-----SLNGVDRKLLSAVEKMLVOKGAPNGIIV----- 48
DB 32 IKILNDSNSLVTLDVINGKFSYIGNLNGAEKKGEGVEEMNLKIIDSNGIAPGV 91
QY 49 -----VF--GLLLYALA-----ARTSPKVRADSDVFNSSFGERNVVTGDLKVLDCGA 99
DB 92 GSHSHVFGSDRHRHAPMLAGATVNEVQAGGIGFTTN---KTRPASEQLRQDFELTA 148
QY 100 FLTRFTNKLRTFGRTFETAVYDFCIAYKHLFOLN--AAALGIPADSYLADFLG--T 155
DB 149 -----KMLRSGITTLLEKSGYGLNVDAMEMKLVLTENPNPLP---VSATFCGANA 199
QY 156 CPKLS--ELQOSR-----KMPASVYALKTEGAVNTPVSNL 189
DB 200 VPKSTFVQTMTICEELIPKIEDRKNGLKN--VENI 236

RESULT 14

G75201
probable beta-lactamase (EC 3.5.2.6) PAB0087 - Pyrococcus abyssi (strain Orsay)
N:Alternate names: penicillin-binding protein homolog
C:Species: Pyrococcus abyssi
C:date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 05-Nov-1999
C:Accession: G75201

R:Anonymous, Genoscope

submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence; insights into archaeal chromosome struc

A:Reference number: A75001
A:Accession: G75201
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-447 <KAN>
A:Cross-references: GB:A1248283; GB:AL096836; NID:95457433; PIDN:CA849062.1; PID:e151495
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: bpd; PAB0087
C:Keywords: hydrolase

Query Match 8.7%; Score 86; DB 2; Length 447;
Best Local Similarity 26.0%; Pred. No. 3.9;
Matches 40; Conservative 15; Mismatches 41; Indels 58; Gaps 8;

QY 26 KLLSAVEKMLVOKGAPNGIEVFGLLYALAAARTSPKVRADSDVFNSSFGERNV-- 84
DB 5 KLSFIVEKMAERK--VFGISISII-----KQGVVYAKGFGRVVE 44
QY 85 -----VTEGDLKVLDCAPLTFPTN--KLRTFGRTFT 116
DB 45 ARLPSTPRTYIGISITSSFTALAMKLYEBGSL--LDD--PVEKFNILKLRPPGEPT 100
QY 117 EAVVDFCIAYKHLFOLNAAAL--GIPADSYL 148
DB 101 ---VHLLTHSSGIPSLGYABAFIDGMVGGDWL 131

RESULT 15

hypothetical protein b2627 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 01-Mar-2002
C:Accession: T08639; E65041
R:Punkett, G.
submitted to the EMBL Data Library, September 1995
A:Reference number: Z16465
A:Accession: T08639
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-729 <PLU>
A:Cross-references: EMBL:U36840; NID:G1033110; PID:G1033122
A:Experimental source: strain K12, substrain MG1655
R:Blattner, F.R.; Punkett, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co.
A:Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; NCID:97426617; PMID:9278503
A:Accession: E65041
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-729 <BLAT>
A:Cross-references: GB:A600348; GB:U00096; NID:G1788975; PIDN:AAC75675.1; PID:G1788980;
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Map position: 57 min

Query Match 8.6%; Score 85.5; DB 2; Length 729;
Best Local Similarity 22.4%; Pred. No. 8.1;
Matches 41; Conservative 31; Mismatches 64; Indels 47; Gaps 8;

QY 2 ELMSDSNLNLTVDASSLNGVDRKLLSAVEKMLVOKGAPNGIEVFGLLYALAA 59
DB 520 EAMDNFFSLVHHMPTNTFLTKTIAKLRLVPHFTSRNGVVKPTDVIATLAGYLSA 579
QY 60 RTSPKVRADSDVFNSS--SFGERNV--VTEGDLKVLDCAPLTFPTNKLRTFGRT 114
DB 580 ESYSEYLRK---NQIDVARQWISGEKRTLSIALNNDLKI-----TN---TFGIT 623
QY 115 FTAA-----YVDFCIAYK--HKLPLNAAALGIPADSYLAA 150

Fri Apr 9 15:30:08 2004

us-09-613-486-15.rpr

Page 6

Db	624	LPKVLSTMEDVVKHAYKRGIRSKVDYTHVYKLA	FESEFHLPPGVNALEIEIGIPQTLHRLV	683
Qy	151	DFL	153	
Db	684	DLL	686	

Search completed: April 5, 2004, 17:54:42
Job time : 22 secs


```

QY 27 LLSAEVKMLVOKGAPNEGIEVFGLLVLAARTSPKVRADSPVFSNGERNVYV 86
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 339 LIQENTERILL-----PHEQPCIIIFERLSLFGCKVTS-----DELINEEDNVSARLVF 388
QY 87 TEGDLKKVLDGCAPLIRFNKLT---FQPTFEAVYDPCIAVYKHLPOLNAAELGIP 142
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 389 ITWMI-----GADVLARKFSEKKITANLDIRSNLEVPVKIYPAHELLISLVSHNLSLD 443
QY 143 AEDSVLADFLGTCPKLSELQGRKMFASVYALKTEGGVNTFVPSNRL 192
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 444 -----LPLDWMERCYKLRDLISNN-----LRSFNG---KPTLRQL 478

RESULT 3
AXP_YARLI STANDARD; PRT; 397 AA.
ID AC Q93389;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Acid extracellular protease precursor (EC 3.4.23.-).
GN AXP.
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxId=4952;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=148;
RX MEDLINE=97039661; PubMed=8885407;
RA Young T.W., Madson A., Glover D.J., Quincey R.V., Butlin M.J.,
RA Kamei E.A.;
RT "The extracellular acid protease gene of Yarrowia lipolytica:
RT sequence and pH-regulated transcription.";
RL Microbiology 142:2913-2921(1996).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to peptidase family A1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X97068; CA65778.1; -.
DR HSBP; P07339; LYB.
DR MEROPS; A01.036; -.
DR InterPro; IPR001969; Asprotease AS.
DR InterPro; IPR009007; Pept A acid.
DR InterPro; IPR001461; Peptidase_A1.
DR Pfam; PF00026; asp.1.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
KW Hydrolyase; Aspartyl protease; Glycoprotein; Zymogen; signal.
FT SIGNAL 1
FT PROPEP 18
FT CHAIN 17
FT ACT_SITE 77
FT ACT_SITE 264
FT DISULFID 93
FT DISULFID 343
FT CARBOHYD 88
FT CARBOHYD 310
FT CARBOHYD 314
SQ SEQUENCE 397 AA; 42081 MW; BD678814B8984F CRC64;

Query Match 9.0%; Score 89.5; DB 1; Length 397;
Best Local Similarity 21.9%; Pred. No. 1.1;
Matches 46; Conservative 37; Mismatches 86; Indels 41; Gaps 8;

```

```

QY 15 TDASINGVDKKLLSAEVEKMLVOKG-APNEGIEVFGLLVLAARTSPKVRAD--S 71
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 136 TDNAVIGA-----SAPGFRGVNSGDLSSGFSVVFQVGNVSDASTSISQLOKSGEIS 190
QY 72 DVIFSNFSGERNVVTTEGDLKKVLDGCAPLIRFNKLTFRPTE-AYVDFCIAYKHL 130
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 191 RNLVGNMFSQANLACTSNDSSEITFGAINTGRVYGSGLKTIIRVAVTQGGYQVFSVSAQKF 250
QY 131 POL-----NAAAEGLGIPAEVSYIAD---FLG-----TCPKSELQ 163
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 251 GDVDFDNDVILDSGTTMTYLSKSDYNAFLGGLDLDLITLSDXSGWGHGVPCESENKIN 310
QY 164 -----QSRKMFASVYALKTEGGVNTFVPSN 188
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 311 FTYNFGKEITVTGHLAIPGNAVNSVDS 340

RESULT 4
PCK_BACSU STANDARD; PRT; 527 AA.
ID AC P54418; O34304;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phosphoenolpyruvate carboxylase [ATP] (EC 4.1.1.49) (PEP
DE carboxylase) (Phosphoenolpyruvate carboxylase) (PEPCK).
GN PCKA OR PCK OR BS030560.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxId=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98048467; PubMed=9387221;
RA Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;
RT "Sequencing and functional annotation of the Bacillus subtilis genes
RT in the 200 kb rmb-dnaB region.";
RL Microbiology 143:3431-3441(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogatawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Berteiro M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Bronkist S., Brunsch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
RA Dentzor F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Eutlian K.D., Evington J., Fabeet C., Ferrati E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golligly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Haelech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Kleaer-Blanchard M., Klein C.,
RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinis S., Lauder J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Paro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Priescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche E., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B.,
RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni K.,
RA Tosiato V., Uchiyama S., Vandenbol M., Vannier F., Vassaretti A.,
RA Viari A., Wambut R., Yamamoto H., Yamae K., Yasunoto K., Yata K.,
RA Winters P., Wipat A., Yoshikawa H.F., Zimstein E., Yoshikawa H., Danchin A.,
RA Yoshida K., Yoshikawa H.F., Zimstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).

```

[3]
 RN SEQUENCE OF 1-165 FROM N.A.
 RP STRAIN=168 / PY79;
 RC MEDLINE=96345628; PubMed=8755891;
 RA Youm R., Perkins J.B., Howitt C.L., Pero J.,
 RT "Cloning and characterization of the metE gene encoding S-adenosylmethionine synthetase from *Bacillus subtilis*.";
 RL J. Bacteriol. 178:4604-4610(1996).
 CC -1- CATALYTIC ACTIVITY: ATP + oxaloacetate = ADP + phosphoenolpyruvate + CO(2).
 CC -1- PATHWAY: Rate-limiting gluconeogenic enzyme.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: Belongs to the phosphoenolpyruvate carboxykinase (ATP) family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to license@ebi.ac.uk).
 CC -----
 CC EMBL: AF008220; AAC00377.1; -
 DR EMBL: Z99119; CAB15034.1; -
 DR EMBL: U52812; AAB17065.1; -
 DR PIR: F69673; P69673.
 DR HSSP: P22259; 1AYL.
 DR Subtilist: Bg1841; pckA.
 DR HAMAP: MF_00453; -; 1.
 DR InterPro: IPR001272; PEPCK_ATP.
 DR Pfam: PF01293; PEPCK_ATP; 1.
 DR ProDom: PD004723; PEPCK_ATP; 1.
 DR TIGRFAMs: TIGR00224; pckA; 1.
 DR PROSITE: PS00532; PEPCK_ATP; 1.
 DR Gluconeogenesis; Lysase; Decarboxylase; ATP-binding; Complete proteome.
 KM NP_BIND 233 240 ATP (By similarity).
 FT CONFLICT 10 10 L -> S (IN REF. 3).
 SQ SEQUENCE 527 AA; 58300 MW; E51EEC802D1E666 CRC64;
 Query Match 8.8%; Score 87; DB 1; Length 527;
 Best Local Similarity 26.9%; Pred. No. 2.6;
 Matches 47; Conservative 24; Mismatches 70; Indels 34; Gaps 9;
 QY 20 LNSGVKLLSAEVEKMLV--QKAPNDEIVVFGLLIYLAATTSBPQVADSDVIFS 76
 DB 234 LSGTGTLLSADADRKLTIGDEHGWSDYGFNIIEG-GCYAKCIHISEEK---EPQIFNA 288
 QY 77 NSPGE--RNVVTEGDLKKVLDGCAPLTFRTNKLRTFGRTTEAVVDFCIAYK-HKLPOL 133
 DB 289 IRFGVLENVAVDDE-----TREANVYDSSFTEENTRA-----AYPIHMINNI 330
 QY 134 NAAAEGLIPADSVYADFLGTCPKLSLQOSRKM--FASNYALK--TEGGVYN 183
 DB 331 VTPSWAGHPSAIVFLTADAFGLTLPISKLTKEQVYHPLSGYTSKLACTGERTGVS 385
 RESULT 5
 YFUK_ECOLI
 ID YFUK_ECOLI STANDARD; PRT; 729 AA.
 AC P52126;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein yfjK.
 GN YFUK OR B2627
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.
 OC NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MC1655;

RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. II, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-VIDE J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to license@ebi.ac.uk).
 CC -----
 CC EMBL: U36840; AAA79796.1; -
 DR EMBL: AE000348; AAC75675.1; -
 DR PIR: T08639; T08639.
 DR EcoGene: EG13197; yfjK.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR001650; Helicase_C.
 DR Pfam: PF00271; Helicase_C; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00490; HELIC_C; 1.
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 729 AA; 83061 MW; C865BCD1B36PFAF CRC64;
 Query Match 8.6%; Score 85.5; DB 1; Length 729;
 Best Local Similarity 22.4%; Pred. No. 5.2;
 Matches 41; Conservative 31; Mismatches 64; Indels 47; Gaps 8;
 QY 2 ELMSDSNLSNLYT--TDASSLNGVKKLLSAEVEKMLVQKAPNDEIVVFGLLIYLA 59
 DB 520 EMMDDNEFSLVFHMTPTNFKTPAKTIARLVPTFSRNGVYKPTVMIAKLGYISA 579
 QY 60 RTSPKPVQVADSDVIFSN---SFGERNV--VTEGDLKKVLDGCAPLTFRTNKLRTFGRT 114
 DB 580 ESYSEYLRK---NQIDYAAQWISSEKRTLSALNNDLKI-----TN---TFGYT 623
 QY 115 FTBA-----YDFCIAYK-HKLPOLNAAAEGLIPADSVYLA 150
 DB 624 LPKVSLMEDVYKHAVKRGRSKYDTHVKLAFESFHPGVNALLEIGLPIQLHPLV 683
 QY 151 DFL 153
 DB 684 DFL 686
 RESULT 6
 PPCK_RHOPA
 ID PPCK_RHOPA STANDARD; PRT; 537 AA.
 AC Q9ZNH4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Phosphoenolpyruvate carboxykinase [ATP] (EC 4.1.1.49) (PEP carboxykinase) (Phosphoenolpyruvate carboxylase) (PEPCK).
 GN PCKA.
 OS Rhodospseudomonas palustris.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Bradyrhizobiaceae; Rhodospseudomonas.
 OC NCBI_TaxID=1076;
 RN [1]
 RP SEQUENCE FROM N.A., SEQUENCE OF 1-12, AND DEVELOPMENTAL STAGE.
 RC STRAIN=7;
 RX MEDLINE=99235744; PubMed=10217755;
 RA Inui M., Nakata K., Roh J.H., Zahn K., Yukawa H.;
 RT "Molecular and functional characterization of the Rhodospseudomonas palustris no. 7 phosphoenolpyruvate carboxykinase gene";
 RL J. Bacteriol. 181:2685-2696(1999).
 CC -1- CATALYTIC ACTIVITY: ATP + oxaloacetate = ADP + phosphoenolpyruvate + CO(2).


```

CC -1- PATHWAY: Rate-limiting gluconeogenic enzyme.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED AT HIGH LEVELS DURING LOG PHASE
CC WITH 10-20 FOLD REDUCTION AT ONSET OF STATIONARY PHASE.
CC -1- SIMILARITY: Belongs to the phosphoenolpyruvate carboxykinase (ATP)
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB015618; BAA34956.1; -.
DR HSSP; P22259; 1A02.
DR HAMAP; MF_00453; -.
DR InterPro; IPR001272; PEPCK_ATP.
DR Pfam; PF01293; PEPCK_ATP; 1.
DR ProDom; PD004723; PEPCK_ATP; 1.
DR TIGRFAMs; TIGR00224; pckA; 1.
DR PROSITE; PS00532; PEPCK_ATP; FALSE NEG.
KW Gluconeogenesis; Lyase; Decarboxylase; ATP-binding.
FT NP BIND 236 243 ATP (By similarity).
SQ SEQUENCE 537 AA; 58773 MW; 98F0E1CED1C9EA6 CRC64;

Query Match 8.5%; Score 84; DB 1; Length 537;
Best Local Similarity 26.7%; Pred. No. 4.9;
Matches 48; Conservative 21; Mismatches 73; Indels 38; Gaps 10;

QY 20 LNVGDKLLSAEVEMLVQKAPNEGJFEVFLD-LYALAARTS--PKVGRADSDVI 74
DB 237 LSGGKTLSDPNRTLIGDEHGMGKDVNPEFGCCAKIKLSANEPRIYA----- 291
QY 75 FSNSEGE--RNVTYEGDLKYLQCAPLRFYTKLRTFGFTTEAYVDFCIANK-HKLP 131
DB 292 -STRFGAVLENNVLCIDIRKPDFDGSK---TENTTS-----AYPLESIP 332
QY 132 QLNAAHEGIPADESYLAADPLGCPKSEIQQSKM--FASWYALK---TEGGVNA-TP 185
DB 333 NASLTGRAGQPRNVMALADAFGWMPYIAKLTPQAMTHFSITAKYAGIERGVTPPT 392

RESULT 7
TILD_BUCAI STANDARD; PRT; 483 AA.
ID TILD_BUCAI
AC P57478
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DR TILD protein homolog.
GN Buchnera sp.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=116093;
RN SEQUENCE FROM N.A.
RP STRAIN=TOKYO 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS."
RL Nature 407:81-86(2000).
CC -1- SIMILARITY: Belongs to the tild/pmba family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP001119; BAB3101.1; -.
DR InterPro; IPR002510; Pepsidase_U62.
DR Pfam; PF01523; Pmba_T1d; 1.
KW Complete proteome.
SQ SEQUENCE 483 AA; 52977 MW; A05CE98518720EBA CRC64;

Query Match 8.4%; Score 83.5; DB 1; Length 483;
Best Local Similarity 20.5%; Pred. No. 4.8;
Matches 32; Conservative 26; Mismatches 53; Indels 45; Gaps 5;

QY 65 KYGRADSDVIFNSFGERNVVTEGDLKKVD-----GCAPLTR 103
DB 305 KNGRGLSIDDEGCTGGKNILKNGILKKYQDKNAALMGVKSNGRRESYCLPMPR 364
QY 104 FTKLRTFGRTTEAYVDFCIAYKHLPOLNAAELGIPADSYLAADPLGCPKLSIQ 163
DB 365 MTN-----TYMLSGSKLDDIISVDYGI-----YAVNFGS---GQVDIT 401
QY 164 QSRKWPASWYALKTEGGVNVFVSNLROLGR-REYV 198
DB 402 SGRFVSTSEAYLKNKIVIPDKYITTLIGSGLVW 437

RESULT 8
ARGI_LEPIN STANDARD; PRT; 385 AA.
ID ARG1_LEPIN
AC Q8EY78;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Arginine biosynthesis bifunctional protein argi [includes: Glutamate
DE N-acetyltransferase (EC 2.3.1.35) (Ornithine acetyltransferase)
DE (Ornithine transferase) (OATase); Amino-acid acetyltransferase
DE (EC 2.3.1.1) (N-acetylglutamate synthase) (AGS)] [contains: Arginine
DE biosynthesis bifunctional protein argi alpha chain; Arginine
DE biosynthesis bifunctional protein argi beta chain].
GN ARG1 OR LA4105.
OS Leptospira interrogans.
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX NCBI_TaxID=173;
RN SEQUENCE FROM N.A.
RP STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
RX MEDLINE=22598143; PubMed=12712204;
RA Ren S.-X., Fu G., Jiang X.-G., Zeng R., Miao Y.-G., Xu H.,
RA Zhang Y.-X., Xiong H., Lu G., Lu L.-F., Jiang H.-Q., Jia J., Tu Y.-F.,
RA Jiang J.-X., Gu W.-Y., Zhang Y.-Q., Cai Z., Sheng H.-H., Yin H.-F.,
RA Yao Z.-J., Shen Y., Qiang B.-Q., Xia Q.-C., Guo X.-K., Danchin A.,
RA Saint Girons I., Somerville R.L., Wen Y.-M., Shi W.-H., Chen Z.,
RA Xu J.-G., Zhao G.-P.;
RT "Unique physiological and pathogenic features of Leptospira
RT interrogans revealed by whole-genome sequencing."
RL Nature 422:888-893(2003).
CC -1- FUNCTION: Catalyzes two activities which are involved in the
CC cyclic version of arginine biosynthesis: the synthesis of
CC acetylglutamate from glutamate and acetyl-CoA, and of ornithine by
CC transacetylation between acetylornithine and glutamate (By
CC similarity).
CC -1- CATALYTIC ACTIVITY: N(2)-acetyl-L-ornithine + L-glutamate = L-
CC ornithine + N-acetyl-L-glutamate.
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + L-glutamate = CoA + N-acetyl-L-
CC glutamate.
CC -1- PATHWAY: Arginine biosynthesis; first step.
CC -1- PATHWAY: Arginine biosynthesis; fifth step.
CC -1- SUBUNIT: Heterotetramer of two alpha and two beta chains (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- MISCELLANEOUS: Some bacteria possess a monofunctional argi, i.e.,
CC capable of catalyzing only the fifth step of the arginine

```

```

CC biosynthetic pathway.
CC -1- SIMILARITY: Belongs to the argJ family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AEO11564; AANS1303.1; -.
CC HAMAP; MF_01106; -; 1.
CC InterPro; IPR002813; ArgJ.
CC Pfam; PF01960; ArgJ; 1.
CC ProDom; PD004193; ArgJ; 1.
CC TIGRFAMs; TIGR00120; ArgJ; 1.
CC Arginine biosynthesis; Multifunctional enzyme; Transferase;
CC Acyltransferase; Complete proteome.
CC CHAIN 1 178 ARGININE BIOSYNTHESIS BIFUNCTIONAL
CC CHAIN 179 385 PROTEIN ARGJ ALPHA CHAIN (BY SIMILARITY).
CC CHAIN 179 385 ARGININE BIOSYNTHESIS BIFUNCTIONAL
CC SITE 178 179 PROTEIN ARGJ BETA CHAIN (BY SIMILARITY).
CC SITE 178 179 CLEAVAGE (NONHYDROLYTIC) (BY SIMILARITY).
CC SEQUENCE 385 AA; 4145 MW; 8A525296D0AD7C2F CRC64;

Query Match 8.4%; Score 83; DB 1; Length 385;
Best Local Similarity 21.3%; Pred. No. 4.1;
Matches 29; Conservative 23; Mismatches 44; Indels 40; Gaps 4;

OY 7 SNLSNLYTTDASS---LNGVDKLLSAEVENKLVOKAPN-----EG 45
DB 262 TKLELTISGAKSAQAARKIGKSLNSPLVTATYGGDPWGRLLIMAVGVKPEPPIFEG 321
OY 46 IEVVEGGLLVLAARTSPKQVADSPVIFNSGFERVVVTEGDLKKVLDGCAPLFRFT 105
DB 322 LQIVFG---TLPVKEANPELTKLSEVLYKNTLISLVVLTAVGTIMKTFMGC----- 370

OY 106 NKLRTFGRTFEAYVD 121
DB 371 -----DTEKYIE 378

RESULT 9
PCK AGRTS STANDARD; PRT; 536 AA.
ID PCK AGRTS STANDARD; PRT; 536 AA.
AC Q8BU94;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Phosphoenolpyruvate carboxykinase (ATP) (EC 4.1.1.49) (PEP
DE carboxykinase) (phosphoenolpyruvate carboxylase) (PEPCK).
GN PCCK OR ATU0035 OR ACR_C56.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Secubal U.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendinning J., Deacher G., Gillet M., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Sasmithmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Petry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RT Science 294:2317-2323 (2001).

```

```

RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RX Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RX Querillo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RX Hounzel K., Gordon J., Vaudin M., Iarcoux O., Bep A., Liu F.,
RX Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Marxiz B.,
RX Flanagan C., Crowell C., Garsen J., Lomo C., Sear C., Strub G.,
RX Cleo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RT Science 294:2323-2328 (2001).
CC -1- CATALYTIC ACTIVITY: ATP + oxaloacetate = ADP + phosphoenolpyruvate
CC + CO(2).
CC -1- PATHWAY: Rate-limiting gluconeogenic enzyme.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the phosphoenolpyruvate carboxykinase (ATP)
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AEO08978; AAL4106.1; ALT_INIT.
CC EMBL; AEO07946; AAK8585.1; ALT_INIT.
CC HAMAP; MF_00453; -; 1.
CC InterPro; IPR001272; PEPCK ATP.
CC Pfam; PF01293; PEPCK ATP; 1.
CC ProDom; PD004723; PEPCK ATP; 1.
CC TIGRFAMs; TIGR00224; PCKA; 1.
CC PROSITE; PS00532; PEPCK ATP; 1.
CC Gluconeogenesis; Lyase; Decarboxylase; ATP-binding; Complete proteome.
CC FM BIND 236 243 ATP (BY SIMILARITY).
CC SEQUENCE 536 AA; 57882 MW; 6F90B8F48C86A5A CRC64;

Query Match 8.4%; Score 83; DB 1; Length 536;
Best Local Similarity 26.9%; Pred. No. 6.1;
Matches 47; Conservative 18; Mismatches 72; Indels 38; Gaps 9;

OY 20 LNGVDKLLSAEVENKLVOKAPNNGEIEVFGLL--LYLAARTS---PKVQADSPVI 74
DB 237 LSGTGKTLISADPARLILGDDHGWGHEGIFVEGCGYAKAIKLSSEAEPEIYAA----- 291
OY 75 FNSNFGEE--RNVYVREGDKLVLDGCAPLRTTNLRTFRFTFAVYDFCIAYK-HKLP 131
DB 292 -TNRFTGVLENVLDSESRPFDNDN--SLTENTRS-----AYLHRTIP 331

OY 132 QLNAAELGIPADSYLAADFLGTCPKXSEL--QOSRKNFASMYLK---TEGGV 181
DB 332 NASETIGAHPTIIMLTADAFGVLPPIARLPPEQMMYHFLSGYAKVAGTERGV 386

RESULT 10
ACKA THETN STANDARD; PRT; 401 AA.
ID ACKA THETN STANDARD; PRT; 401 AA.
AC OGR9V4;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Acetate kinase (EC 2.7.2.1) (Acetate kinase).
GN ACKA OR TTE1481.
OS Thermanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermanaerobacteriales;
OC Thermanaerobacteriaceae; Thermanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=MB4 / JCM 11007;
RA MEDLINE=21992816; PubMed=11997336;

```

```

RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.,
RT "A complete sequence of T. tengcongensis genome."
RL Genome Res. 12:689-700(2002).
CC -1- CATALYTIC ACTIVITY: ATP + acetate = ADP + acetyl phosphate.
CC -1- PATHWAY: Conversion of acetate to acetyl-CoA, first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the acetate kinase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE013105; AAM24700.1; -.
DR HAMAP; MF_00020; -. 1.
DR InterPro; IPR000890; Acetate_kin.
DR InterPro; IPR004372; AckA.
DR Pfam; PF00871; Acetate_kinase; 1.
DR PRINTS; PR00471; ACETATEKINASE.
DR TIGRFAMs; TIGR00016; ackA; 1.
DR PROSITE; PS01075; ACETATE_KINASE_1; 1.
DR PROSITE; PS01075; ACETATE_KINASE_2; 1.
DR TRANSFERASE; Kinase; Complete proteome.
SQ SEQUENCE 401 AA; 44361 MW; PD71C642F1E01950 CRC64;

Query Match 8.2%; Score 81.5; DB 1; Length 401;
Best Local Similarity 19.3%; Pred. No. 5.9;
Matches 40; Conservative 35; Mismatches 87; Indels 45; Gaps 7;

QY 22 GVDKLLSAAY---EKMLYOKGAPN--EGLEVFGLL-YALAAARTSPKVCADSDVIF 75
DB 35 GINSLSLTHGKGEKVKYKIDMKMKKEALQVLEVLVDKEIGVYDKMEIDAVGRVHA 94
QY 76 SNSFGERNVVTGDLKKVLGGC---APLTRFTN-----KLRFGFTTE 117
DB 95 GGEVFTDSVLIDDEVIKK-LEDICIDLAPLNHPANIEGKAKCOQIMPSVPVAVFDIAFQ 153
QY 118 AYVDFCLAY-----KHKLPQNAAEGLPREDSYLAADPLGTCP 157
DB 154 TMPYAVYIPYEPYVKKHRIKRRYGFHGTSHKRYSNPAABITLGRPIELKIVTCHLNGA 213
QY 158 KLSLQOSRKMFASMYALKTEGGVYNT 184
DB 214 SITAVKXGKSIDISMGFPLEGLAMGT 240

RESULT 11
LON2 BORBU STANDARD; PRT; 813 AA.
ID LON2 BORBU
AC OS1558;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ATP-dependent protease La homolog (EC 3.4.21.-).
GN BB0613
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A. / B31;
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Iachigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kesteven A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.P., Gocayne J.D., Weidman J.,
RA Uitterback L., McQuay L., McDonald L., Artiach P., Bowman C.,

```

```

RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT burgdorferi."
RL Nature 390:580-586(1997).
CC -1- SIMILARITY: Belongs to peptidase family S16.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001162; AAC69662.1; -.
DR PIR; D70176; D70176.
DR MEROPS; S16.DPW; -.
DR TIGR; B0613; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003959; AAA_ATPase_cent.
DR InterPro; IPR008269; Pept_S16_C.
DR InterPro; IPR004815; Pept_S16_Ion.
DR InterPro; IPR003111; Pept_S16_N.
DR InterPro; IPR008268; Peptid_S16_AS.
DR InterPro; IPR01984; Peptidase_S16.
DR Pfam; PF00004; AAA; 1.
DR Pfam; PF02190; LON; 1.
DR Pfam; PF05362; Lon_C1.
DR PRINTS; PRO0830; ENDOLAPTASE.
DR SMART; SM00382; AAA; 1.
DR SMART; SM00464; LON; 1.
DR TIGRFAMs; TIGR00763; lon; 1.
DR PROSITE; PS01046; LON_SER; 1.
KW Hydrolase; Serine protease; ATP-binding; Complete proteome.
FT NP BIND 369 376
FT ACT SITE 719 719 BY SIMILARITY.
SQ SEQUENCE 813 AA; 92312 MW; 5129AA1498CD5D0F6 CRC64;

Query Match 8.2%; Score 81.5; DB 1; Length 813;
Best Local Similarity 31.9%; Pred. No. 14;
Matches 37; Conservative 19; Mismatches 43; Indels 17; Gaps 6;

QY 8 NLSN-LVITDASLNGVDKLLSAVERKLYOKGAPNGIEVFGLLVLAARTSPKV 66
DB 475 DLNVLFTTANSLNGMKRPLID-RMEIITKYGSIYIKLAIKFLIPSIKXSFIDKV 533
QY 67 Q-RADSDVIFSNFSGERNVVTG--DLKYLIDGCAPLTRFTNKLRTFGRTTAY 119
DB 534 YRIEDDYIFNLI---RNVTWESGVRLKRVL-----TNLIRLVRELLEY 577

RESULT 12
SPSY MOUSE STANDARD; PRT; 366 AA.
ID SPSY MOUSE
AC P97355;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Spermine synthase (EC 2.5.1.22) (Spermidine aminopropyltransferase)
DE (SPMSY).
GN SMS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RP Strom T.M., Francis F., Lorenz B., Boedich A., Boone M.J.,
RA Lehrsich H., Mellingner T.,
RT "pex gene deletions in Gy and Hyp mice provide mouse models for
RT X-linked hypophosphatemia."
RT Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.

```

RN [2] SEQUENCE FROM N.A.
 RC TISSUE=Diaphragm.
 RA Nitraren K., Korhonen V., Janne J.,
 RT "Nucleotide sequence of mouse spermidine aminopropyltransferase
 CDNA."
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Olfactory epithelium;
 RX MEDLINE=2238257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Altschuler R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Alekshun S.F., Zeeberg B., Buetow K.H., Scheef C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Dichtenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stjepanec M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci F., Pirange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Wuzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting W., Madan A., Young A.C., Shcherbko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butcherfield V.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16839-16903(2002).
 RN [4]
 RP SEQUENCE OF 316-366 FROM N.A.
 RX MEDLINE=9813337; PubMed=9467015;
 RA Lorenz B., Francis F., Gempel K., Boeddrich A., Josten M., Schmahl W.,
 RA Schmidt J., Lebrach H., Meitinger T., Strom T.M.;
 RT "Spermine deficiency in Gy mice caused by deletion of the spermine
 synthase gene."
 RL Hum. Mol. Genet. 7:541-547(1998).
 CC -I- CATALYTIC ACTIVITY: S-adenosylmethioninamide + spermidine = 5'-
 methylthioadenosine + spermine.
 CC -I- PATHWAY: Biosynthesis of spermine from spermidine.
 CC -I- SIMILARITY: Belongs to the spermidine/spermine synthase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL, Y09419; CA70573.1; -
 DR EMBL, AF031486; AAB8631.1; -
 DR EMBL, BC046623; AAB46623.1; -
 DR EMBL, AJ000093; CAA03918.1; -
 DR EMBL, AJ000087; CAA03918.1; -
 DR EMBL, AJ000088; CAA03918.1; JOINED.
 DR EMBL, AJ000089; CAA03918.1; JOINED.
 DR EMBL, AJ000090; CAA03918.1; JOINED.
 DR EMBL, AJ000091; CAA03918.1; JOINED.
 DR EMBL, AJ000092; CAA03918.1; JOINED.
 DR MGD; MG:109430; Sms.
 DR GO; GO:0008215; P:spermine metabolism; IMP.
 DR InterPro; IPR01045; Spermine synthase.
 DR Pfam; PF01564; Spermine synth. 1.
 DR PROSITE; PS00097; CARBAMOYLTRANSFERASE; 1.
 KW TRANSFERASE; 192
 FT DOMAIN 192 232
 FT BINDING TO DECARBOXYLATED SAM
 FT (POTENTIAL).
 FT
 SC SEQUENCE 366 AA; 4133 MW; D549F319F51C43C5 CRC64;
 Query Match 8.1%; Score 80; DB 1; Length 366;

Best Local Similarity 22.9%; Pred. No. 7.2;
 Matches 52; Conservative 34; Mismatches 65; Indels 76; Gaps 13;
 QY 24 DKLLSAVEKMLVQKAPNGIEV-----FGLLVLAARTPKVQRADSDVIFNSF 79
 DB 128 DERLVEYDIEVDVDESPYQNIKILSKQFNTL-----ILSGDVLAESDLATYAI 181
 QY 80 -----GERNVVTEGD-----LK-----KVLDCAPLRFNTKLT 110
 DB 182 MSGKEDVTGQDVLILGGDGGILCEIVKLPKVTWVEIDQWYDCKKYM-----RT 236
 QY 111 FERTTT-----EAYVDFCI-----AYHKLPQNLAAELGIPADSVLADF 152
 DB 237 CDVDLNTKGDYQYLIDCIPVLKMYAKEGREFDYVNDLVAVISTSPEDS--TWDF 294
 QY 153 LGTCPLS--ELQGRKMFASWYALKTEGVN--TPVSNL--RQLGR 194
 DB 295 LRLIDLSKVKLKQDKYF-----TQNCVNLTEALSLVEQLGR 334
 RESULT 13
 ID OTC SYN3 STANDARD; PRT; 308 AA.
 AC 055497;
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Ornithine carbamoyltransferase (EC 2.1.3.3) (OTCcase).
 GN ARGF OR SL10902.
 OS Synchocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
 CX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96127529; PubMed=8590279;
 RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
 RA Sugita M., Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synchocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
 RT region from map positions 64% to 92% of the genome."
 RL DNA Res. 2:153-166(1995).
 CC -I- CATALYTIC ACTIVITY: Carbamoyl phosphate + L-ornithine = phosphate
 + L-citrulline.
 CC -I- PATHWAY: Arginine biosynthesis; sixth step.
 CC -I- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -I- SIMILARITY: Belongs to the ATCase/OTCase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; D6406; BAA10847.1; -
 DR PIR; S76000; S76000.
 DR HSSP; P04391; IAKM.
 DR HAMAP; MF 01109; -; 1.
 DR InterPro; IPR006130; Asp/Om Cotranf.
 DR InterPro; IPR002292; Orn carbtransf.
 DR InterPro; IPR006131; OTCcase_O.
 DR InterPro; IPR006132; OTCcase_P.
 DR Pfam; PF00185; OTCcase; 1.
 DR Pfam; PF02729; OTCcase; 1.
 DR PRINTS; PR00100; AOTCase.
 DR TIGRFAMs; TIGR00658; orn carb tr; 1.
 DR PROSITE; PS00097; CARBAMOYLTRANSFERASE; 1.
 DR Arginine biosynthesis; Transferase; Complete proteome.
 KW Arginine biosynthesis; 33
 FT SITE 33 57
 FT IMPORTANT FOR STRUCTURAL INTEGRITY (BY
 FT SIMILARITY).
 FT CARBAMOYLPHOSPHATE BINDING (BY
 FT SIMILARITY).

Query Match: 8.0%; Score 79; DB 1; Length 308;
 Best Local Similarity 24.8%; Pred. No. 7.2; Indels 30; Gaps 7;
 Matches 39; Conservative 25; Mismatches 63; Indels 30; Gaps 7;

FT SITE 104 104 CARBAMOYLPHOSPHATE BINDING (BY
 FT SITE 131 131 CARBAMOYLPHOSPHATE BINDING (BY
 FT SITE 144 144 SIMILARITY).
 FT SITE 144 144 IMPORTANT FOR STRUCTURAL INTEGRITY (BY
 FT SITE 264 267 ORNITHINE BINDING (BY SIMILARITY).
 SQ SEQUENCE 308 AA; 33616 MW; C6CD2E498592CFA CRC64;

Query Match: 8.0%; Score 79; DB 1; Length 308;
 Best Local Similarity 24.8%; Pred. No. 7.2; Indels 30; Gaps 7;
 Matches 39; Conservative 25; Mismatches 63; Indels 30; Gaps 7;

QY 28 LSAVEKMLVQKAP-NEG1-----EVFGLLLYLAARTSPKYQGRADSVIFSNFGE 81
 DB 17 LITEBKSLQLAADKSGVLKPKHCKRIIGLLFYKASTRT--RVSFYAMTQLGGQVLD 73

QY 82 RNVVYTE-GDLKVVLDGCAPIRTFNKL--RTGRRTFAYVDFCIAYKHKLQOLNAAAE 138
 DB 74 LNVSTQVGRGEPIQDTARVLDKYIDILAVRTFKQDTLQTFADHA---KMPITNALSD 128

QY 139 LGIPADSYLAADFLGTCEPKLSELQOSRKMFSAMVAL 175
 DB 129 LHPH-----CQLADLQITKECFGLBGL 152

RESULT 14
 CPKC_AGRU STANDARD; PRT; 422 AA.

AC P2446;
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cytochrome P450-p1nf1, plant-inducible (EC 1.14.-.-).
 GN CYP103 OR PINF1 OR VIRH1.
 OS Agrobacterium tumefaciens.
 OC plasmid pT1A6.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
 OK NCBI_TaxID=358;
 RN [1]
 RX MEDLINE=69213933; PubMed=2708311;
 RA Kameneto R.H., Powell A.T., Akiyoshi D.E., Regier D.A.,
 RA Kersebaumer R.A., Nester E.W., Hawes M.C., Gordon M.P.,
 RT "Nucleotide sequence and analysis of the plant-inducible locus p1nf1
 from Agrobacterium tumefaciens."
 RL J. Bacteriol. 171:2506-2512 (1989).

CC -1- FUNCTION: NOT ESSENTIAL FOR VIRULENCE, BUT MAY BE INVOLVED IN THE
 CC DETOXIFICATION OF PLANT PROTECTIVE AGENTS AT THE SITE OF WOUNDING.
 CC -1- INDUCTION: TRANSCRIPTIONALLY ACTIVATED IN THE PRESENCE OF WOUNDED
 CC PLANT TISSUE AND BY PLANT PHENOLIC COMPOUNDS, SUCH AS
 CC ACETOSYRINONE.
 CC -1- SIMILARITY: Belongs to the cytochrome P450 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).

CC EMBL; M19352; AAA82502.1; -;
 DR PIR; A32306;
 DR InterPro; IPR001128; Cytochrome_P450.
 DR Pfam; PF00067; P450_1.
 DR PRINTS; PR00385; P450.
 DR PROSITE; PS00086; CYTOCHROME_P450_1.
 DR Oxidoreductase; Monooxygenase; Electron transport; Heme; Plasmid.
 KW METAL 369 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
 FT SEQUENCE 422 AA; 47519 MW; 6A9FE4AA9B7E2302 CRC64;

Query Match: 8.0%; Score 79; DB 1; Length 422;
 Best Local Similarity 23.6%; Pred. No. 11;
 Matches 51; Conservative 29; Mismatches 60; Indels 76; Gaps 14;

QY 20 LNVGDKLLSAVEKMLVQKAPNEGIE--VFGLLLYLAARTSPKYQGRADSVIFSN 77
 DB 70 LAGTDPRTQIETELM-----NRGVKAGAVDFI-----DHSMLFSN 107

QY 78 --SFGERNVVYTEGDLKVVLDGCAPI-RTFT-----NKLRTFGRTFAYVDFCIAYK 127
 DB 108 GETHGKRSGLSKAFSPRVEALRPEIAKITECLMDLQKVDF--NFTMYAS----- 159

QY 128 HKLPQLNAAAELEGPAD-----SYLADFLGTCPKSELQ----- 163
 DB 160 -QLPALTIASVLAGPSDTPFTFLVYKVRCLSPSRDEFEIEASALELDQYRSVI 218

QY 164 --QSRKM---FASVYALKT--EGGVNTPVSNLRQL 192
 DB 219 ADSGRMRHDFLSRY-LKAVREAGTL-SPIELIMQL 252

RESULT 15
 PCKC_FUSNN STANDARD; PRT; 527 AA.

AC O8RE12;
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Phosphoenolpyruvate carboxylase (ATP) (EC 4.1.1.49) (PEP
 DE carboxylase) (Phosphoenolpyruvate carboxylase) (PEPCK).
 GN PCKA OR FN1120.
 OS Fuschacterium nucleatum (subsp. nucleatum).
 OC Bacteria; Fusobacteriales; Fusobacteriaceae;
 OC Fuschacterium.
 OK NCBI_TaxID=76856;
 RN [1]
 RX STRAIN=ATCC 25586;
 RX MEDLINE=2186394; PubMed=11889109;
 RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
 RA Bhattacharya A., Barman A., Gardner W., Grechkin G., Zhu L.,
 RA Vasileva O., Chu L., Kogan Y., Chaga O., Goldsman E., Berna A.,
 RA Larsen N., D'Souza W., Malinas T., Pusch G., Haselkorn R.,
 RA Fontein M., Kyrides N., Overbeek R.,
 RT "Genome sequence and analysis of the oral bacterium Fuschacterium
 RT nucleatum strain ATCC 25586."
 RL J. Bacteriol. 184:2005-2018 (2002).

CC -1- CATALYTIC ACTIVITY: ATP + oxaloacetate = ADP + phosphoenolpyruvate
 CC + CO(2).
 CC -1- PATHWAY: Rate-limiting gluconeogenic enzyme.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: Belongs to the phosphoenolpyruvate carboxylase (ATP)
 CC family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).

CC EMBL; AE010616; AAL95316.1; -;
 DR HAMAP; MF_00453; 1.
 DR InterPro; IPR001272; PEPCK_ATP.
 DR Pfam; PF01293; PEPCK_ATP_1.
 DR ProDom; PD004723; PEPCK_ATP_1.
 DR TIGRFAMs; TIGR00224; pckA_1.
 DR PROSITE; PS00532; PEPCK_ATP_1.
 DR Gluconogenesis; Lyase; Decarboxylase; ATP-binding; Complete proteome.
 KW NP_BIND 230 ATP (BY SIMILARITY)
 FT SEQUENCE 527 AA; 59055 MW; 275849FD254AC01 CRC64;

Fri Apr 9 15:30:08 2004

us-09-613-486-15.rsp

Page 10

```

Query Match      8.0%; Score 79; DB 1; Length 527;
Best Local Similarity 26.6%; Pred. Nc. 14;
Matches 46; Conservative 26; Mismatches 67; Indels 34; Gaps 10;

QY      20  LNVGDYKLLSLASVEKMLV---QKQAPNKGIEVVFSLLYLAARTSPKRVQADSDVIES 76
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd      21  LSGTGTLLTSDAPNKKLTIGDDEHGCMDCKGIFNFEQ---GCYAKCNINK-EESSEIERYA 285
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      77  NSFGE--RNVVYVTEBDLKKVLDGCAPLRFTYMKLRTPGRTF--EAYVDFCAVYKHKPQL 133
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd      286  IKFGSLVENVYVD-----PIRR---KIQYEDASTENTRVGPIIDY---IPNA 327
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      134  NAAAEELGPAEDSYLAADFLGTCPKLSIEHQSRKM--FASWYALK---TEGCV 181
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd      328  ELISVGIGLPKVIIFILADSPGVGLPIISRLSGQANVYHHVTGTTALAGTSLGCV 380
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: April 5, 2004, 17:53:11
Job time : 19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2004, 17:50:07 ; Search time 45 Seconds
(Without alignments)

1388.280 Million cell updates/sec

Title: US-09-613-486-15

Sequence: 1 MEIMSDSNLSNLVITDASSL.....GVVNTPVNSLRQLGRREVM 198

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_plant:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriophage:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	991	100.0	198	071213 grapevine 1
2	984	99.3	198	039857 grapevine 1
3	921	92.9	198	08BEP3 grapevine 1
4	297	30.0	204	09Q629 sugar beet
5	296	29.9	204	09Q664 sugar beet
6	295	29.8	204	089910 sugar beet
7	293	29.6	204	003724 sugar beet
8	286	28.9	204	008531 sugar beet
9	265	26.9	215	065858 beet yellow
10	130.5	13.2	223	08JYV1 citrus tris
11	130.5	13.1	223	09Q6C4 citrus tris
12	125.5	12.7	222	0990F0 citrus tris
13	125.5	12.7	223	0990P2 citrus tris
14	125.5	12.7	223	09Q6G5 citrus tris
15	125	12.6	223	08B3T3 citrus tris
16	124.5	12.6	223	09Q6G2 citrus tris

17	123.5	12.5	223	12	09E7M2	09E7M2 citrus tris
18	123.5	12.5	223	12	09E7M4	09E7M4 citrus tris
19	123.5	12.5	223	12	09PXT0	09PXT0 citrus tris
20	123.5	12.5	223	12	09Q6G3	09Q6G3 citrus tris
21	123	12.4	212	12	09DMW1	09DMW1 citrus tris
22	123	12.4	223	12	09E7M3	09E7M3 citrus tris
23	120.5	12.2	195	12	08V1P7	08V1P7 citrus tris
24	120.5	12.2	222	12	0990M9	0990M9 citrus tris
25	120.5	12.2	223	12	09Q6G1	09Q6G1 citrus tris
26	120.5	12.2	223	12	P89948	P89948 citrus tris
27	120.5	12.2	223	12	0990P1	0990P1 citrus tris
28	120.5	12.2	223	12	09MTD9	09MTD9 citrus tris
29	120.5	12.2	223	12	091FW4	091FW4 citrus tris
30	120.5	12.2	223	12	006192	006192 citrus tris
31	120.5	12.2	223	12	09PXT3	09PXT3 citrus tris
32	118.5	12.0	223	12	08JYU9	08JYU9 citrus tris
33	117.5	11.9	223	12	09PXS9	09PXS9 citrus tris
34	117.5	11.9	223	12	0991G6	0991G6 citrus tris
35	116	11.7	223	12	09Q6C6	09Q6C6 citrus tris
36	115.5	11.7	195	12	08V1P5	08V1P5 citrus tris
37	115.5	11.7	223	12	08JYV0	08JYV0 citrus tris
38	115	11.6	223	12	09PXT1	09PXT1 citrus tris
39	114.5	11.6	223	12	09DTP8	09DTP8 citrus tris
40	114.5	11.6	223	12	010474	010474 citrus tris
41	114.5	11.6	223	12	08B3R9	08B3R9 citrus tris
42	113.5	11.5	194	12	08V1P6	08V1P6 citrus tris
43	111.5	11.3	223	12	08JYU8	08JYU8 citrus tris
44	110.5	11.2	223	12	09PXT2	09PXT2 citrus tris
45	109	11.0	200	12	099AT6	099AT6 citrus tris

ALIGNMENTS

RESULT 1

ID 071213 PRELIMINARY; PRT; 198 AA.

AC 071213:

DT 01-AUG-1998 (TEMBLrel. 07, Created)

DT 01-AUG-1998 (TEMBLrel. 07, Last sequence update)

DE 01-JUN-2003 (TEMBLrel. 24, Last annotation update)

DE 22 kDa coat protein.

OS Grapevine leafroll-associated virus 2.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;

OC Closterovirus.

OC NCBI_TaxId=64003;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98264507; PubMed=9603345;

RA Zhu H.Y., Ling K.S., Gosczyński D.E., McFerson J.R., Gonsalves D.;

RT "Nucleotide Sequence and Genome Organization of Grapevine Leafroll-1-Associated Virus-2 are Similar to Beet Yellow Virus, the

RT Closterovirus Type Member."

RL J. Gen. Virol. 79:1289-1298(1998).

DR EMBL: AF039204; AAC40861.1; -

DR GO: GO:0019028; C:Viral capsid; IEA.

DR GO: GO:0005198; P:Structural molecule activity; IEA.

DR InterPro: IPR002679; Closter coat.

DR Pfam: PF01785; Closter_coat; 1.

KW Coat protein.

SQ SEQUENCE 198 AA; 21660 MW; 8ED85EF2E1DFB03 CRC64;

Query Match 100.0%; Score 991; DB 12; Length 198;

Best Local Similarity 100.0%; Pred. No. 1.3e-88;

Matches 198; Conservativity 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MEIMSDSNLSNLVITDASSLNGVDKLLSAEYKMLVCGAGNEGIEVFGLLVALAAR	60
DB	1	MEIMSDSNLSNLVITDASSLNGVDKLLSAEYKMLVCGAGNEGIEVFGLLVALAAR	60
QY	61	TTSPKTVQADSDVIFENSGFERNVVVTBGLDKKVIDGCAPLTFPNKLTFRFTTEAVV	120
DB	61	TTSPKTVQADSDVIFENSGFERNVVVTBGLDKKVIDGCAPLTFPNKLTFRFTTEAVV	120

QY 121 DFCIAYVHKHLPOLNAAAEIGIPAEBSYLAADFLGTCPKLSLQSRKMFASMYALKTEGG 180
 DB 121 DFCIAYVHKHLPOLNAAAEIGIPAEBSYLAADFLGTCPKLSLQSRKMFASMYALKTEGG 180
 QY 181 VVNTPVSNLRQLGRREV 198
 DB 181 VVNTPVSNLRQLGRREV 198

RESULT 2

039857 PRELIMINARY: PRT: 198 AA.
 AC 039857
 DT 01-JAN-1998 (Tremblrel. 05, Created)
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
 DE Coat protein.
 OS Grapevine leafroll-associated virus 2.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
 OC Closterovirus.
 OC NCBI_TaxID=64003;
 RN (1)
 RN SEQUENCE FROM N.A.
 RA Abou-Ghanem N.;
 RT "The nucleotide sequence of the 3'terminal region of grapevine
 leafroll associated closterovirus 2,";
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Y14131; CAA74566.1; -
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR002679; Closter_coat.
 DR Pfam; PF01785; Closter_coat; 1.
 SQ SEQUENCE 198 AA; 21630 MW; 15D631E408BD3F01 CRC64;

Query Match 99.3%; Score 984; DB 12; Length 198;
 Best Local Similarity 99.5%; Pred. No. 6,3e-88;
 Matches 197; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MELMSDSNLSNIVITDASSLNGVDKLLSAEYKMLVKGAPNKGIEVFGILLYALAAR 60
 DB 1 MELMSDSNLSNIVITDASSLNGVDKLLSAEYKMLVKGAPNKGIEVFGILLYALAAR 60
 QY 61 TTSPPVQADSDVIFSNFSGERNVVTGGDLKKVLDGCAPLTRFNKLRTRGRTTEAYV 120
 DB 61 TTSPPVQADSDVIFSNFSGERNVVTGGDLKKVLDGCAPLTRFNKLRTRGRTTEAYV 120
 QY 121 DFCIAYVHKHLPOLNAAAEIGIPAEBSYLAADFLGTCPKLSLQSRKMFASMYALKTEGG 180
 DB 121 DFCIAYVHKHLPOLNAAAEIGIPAEBSYLAADFLGTCPKLSLQSRKMFASMYALKTEGG 180
 QY 181 VVNTPVSNLRQLGRREV 198
 DB 181 VVNTPVSNLRQLGRREV 198

RESULT 3

039857 PRELIMINARY: PRT: 198 AA.
 AC 039857
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
 DE Coat protein.
 OS Grapevine rootstock stem lesion associated virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
 OC Closterovirus.
 OC NCBI_TaxID=167634;
 RN (1)
 RN SEQUENCE FROM N.A.
 RA Zhang Y.; Rohhani A.;
 RT "Nucleotide sequence of grapevine rootstock stem lesion associated
 virus,";

RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF314061; AAN63472.1; -
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR002679; Closter_coat.
 DR Pfam; PF01785; Closter_coat; 1.
 SQ SEQUENCE 198 AA; 21620 MW; 370E95A9369734A9 CRC64;

Query Match 92.9%; Score 921; DB 12; Length 198;
 Best Local Similarity 90.9%; Pred. No. 9.1e-82;
 Matches 180; Conservative 10; Mismatches 8; Indels 0; Gaps 0;

QY 1 MELMSDSNLSNIVITDASSLNGVDKLLSAEYKMLVKGAPNKGIEVFGILLYALAAR 60
 DB 1 MELMSDSNLSNIVITDASSLNGVDKLLSAEYKMLVKGAPNKGIEVFGILLYALAAR 60
 QY 61 TTSPPVQADSDVIFSNFSGERNVVTGGDLKKVLDGCAPLTRFNKLRTRGRTTEAYV 120
 DB 61 TTSPPVQADSDVIFSNFSGERNVVTGGDLKKVLDGCAPLTRFNKLRTRGRTTEAYV 120
 QY 121 DFCIAYVHKHLPOLNAAAEIGIPAEBSYLAADFLGTCPKLSLQSRKMFASMYALKTEGG 180
 DB 121 DFCIAYVHKHLPOLNAAAEIGIPAEBSYLAADFLGTCPKLSLQSRKMFASMYALKTEGG 180
 QY 181 VVNTPVSNLRQLGRREV 198
 DB 181 VVNTPVSNLRQLGRREV 198

RESULT 4

090629 PRELIMINARY: PRT: 204 AA.
 AC 090629
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
 DE Major capsid protein.
 OS Sugar beet yellow virus (SBYV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
 OC Closterovirus.
 OC NCBI_TaxID=12161;
 RN (1)
 RN SEQUENCE FROM N.A.
 RP STRAIN=BYV-4;
 RC MEDLINE=20079557; PubMed=10611288;
 RX Peremyshlov V.V.; Hagihara Y.; Dolja V.V.;
 RA "SBYV homolog functions in cell-to-cell movement of a plant virus,";
 RT Proc. Natl. Acad. Sci. U.S.A. 96:14771-14776 (1999).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=BYV-4;
 RC Peremyshlov V.V.; Hagihara Y.; Alzhanova D.; Dolja V.V.;
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF190581; AAF14505.1; -
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR002679; Closter_coat.
 DR Pfam; PF01785; Closter_coat; 1.
 SQ SEQUENCE 204 AA; 22419 MW; FD6337E1D2490D3 CRC64;

Query Match 30.0%; Score 297; DB 12; Length 204;
 Best Local Similarity 37.6%; Pred. No. 8.6e-21;
 Matches 71; Conservative 32; Mismatches 84; Indels 2; Gaps 2;

QY 7 SNLSNIVITDASSLNGVDKLLSAEYKMLVKGAPNKGIEVFGILLYALAARTSPKY 66
 DB 7 SNLSNIVITDASSLNGVDKLLSAEYKMLVKGAPNKGIEVFGILLYALAARTSPKY 66
 QY 11 ATPEVNLADQTCIHGSDCDKLRDFEBCIKLKGVEDKIGLIGLIVSCATIGTSNKV 70
 DB 11 ATPEVNLADQTCIHGSDCDKLRDFEBCIKLKGVEDKIGLIGLIVSCATIGTSNKV 70
 QY 67 QRADSDVIFSNFSGERNVVTGGDLKKVLDGCAPLTRFNKLRTRGRTTEAYVFCIA 125
 DB 67 QRADSDVIFSNFSGERNVVTGGDLKKVLDGCAPLTRFNKLRTRGRTTEAYVFCIA 125
 QY 71 SVQPTSTFIKASFGSGELFTHGELRSFLDSQKLEGGKPKLRCFRTFQKQYISPAKE 130
 DB 71 SVQPTSTFIKASFGSGELFTHGELRSFLDSQKLEGGKPKLRCFRTFQKQYISPAKE 130
 QY 126 YKHKLPOLNAAAEIGIPAEBSYLAADFLGTCPKLSLQSRKMFASMYALKTEGGVNTTP 185
 DB 126 YKHKLPOLNAAAEIGIPAEBSYLAADFLGTCPKLSLQSRKMFASMYALKTEGGVNTTP 185

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: April 7, 2004, 01:15:49 (Search time 2524 Seconds
(Without alignments)
3400.127 Million cell updates/sec

Title: US-09-613-486-15
Perfect score: 991
Sequence: 1 MEIMSDSNLSNLTVDASL.....GGVNTFVSNLRQLGRREV 198

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US09613486/rnatc.05042004.155121.28262/app_query.fasta_1.391
-DB=GenBdb1 -OPMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFM=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09613486 @CGN_1_1_2372.@rnatc.05042004.155121.28262 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEOUTRY -NEG_SCORES=0 -WAIT -DSPELOCK=100 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenBdb1.*
1: gb_da.*
2: gb_neg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pt.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_da.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*

29: em_vl.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rtd.*
36: em_htg_mam.*
37: em_htg_vtc.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	991	100.0	597	6 AR138306	AR138306 Sequence
2	991	100.0	15000	14 AF039204	AF039204 Grapevine
3	991	100.0	15500	6 AR138299	AR138299 Sequence
4	984	99.3	8590	14 GLA44131	V14131 Grapevine 1
5	921	92.9	15527	14 AF314061	AF314061 Grapevine
6	297	30.0	15468	14 AF190581	AF190581 Beet yell
7	296	29.9	612	14 BYU71295	U71295 Beet yellow
8	295	29.8	15468	14 AF056575	AF056575 Beet yell
9	293	29.6	2724	14 BYVC6PG	MS9452 Beet yellow
10	293	29.6	5980	14 BYVMBPA	X73475 Beet yellow
11	286	28.9	6746	14 CLBY73PH	X53462 Beet yellow
12	286	28.9	15480	14 BYVDA	X73476 Beet yellow
13	285	28.8	783	6 A41914	A41914 Sequence 1
14	266.5	26.9	10545	14 BYU51931	U51931 Beet yellow
15	130.5	13.2	672	14 AF501867	AF501867 Citrus tr
16	129.5	13.1	672	14 AF184115	AF184115 Citrus tr
17	125.5	12.7	672	14 AF342894	AF342894 Citrus tr
18	125.5	12.7	672	14 AF184114	AF184114 Citrus tr
19	125.5	12.7	672	14 AF342890	AF342890 Citrus tr
20	125	12.6	1080	14 CTVC0AT	W76485 Citrus tris
21	125	12.6	19293	14 AV170468	AV170468 Citrus tr
22	125	12.6	19296	14 AV340974	AV340974 Citrus tr
23	125	12.6	19296	14 CTU16304	U6304 Citrus tris
24	124.5	12.6	672	14 AF184117	AF184117 Citrus tr
25	123.5	12.5	672	14 AF184116	AF184116 Citrus tr
26	123.5	12.5	672	14 AF220502	AF220502 Citrus tr
27	123.5	12.5	672	14 AF220504	AF220504 Citrus tr
28	123	12.4	672	14 CTR297702	AJ297702 Citrus tr
29	123	12.4	672	14 AF220503	AF220503 Citrus tr
30	120.5	12.2	589	14 AF456771	AF456771 Citrus tr
31	120.5	12.2	669	14 AF342895	AF342895 Citrus tr
32	120.5	12.2	672	14 AF184118	AF184118 Citrus tr
33	120.5	12.2	672	14 AF342891	AF342891 Citrus tr
34	120.5	12.2	672	14 AF342892	AF342892 Citrus tr
35	120.5	12.2	672	14 AF342893	AF342893 Citrus tr
36	120.5	12.2	672	14 CTVC0AT	L12175 Citrus tris
37	120.5	12.2	19226	14 CTU56902	U56902 Citrus tris
38	120.5	12.2	19259	14 AF260651	AF260651 Citrus tr
39	120.5	12.2	19259	14 CTU18440	V18440 Citrus tris
40	118.5	12.0	672	14 AF501869	AF501869 Citrus tr
41	117.5	11.9	587	14 AF220505	AF220505 Citrus tr
42	117.5	11.9	672	14 AF339088	AF339088 Citrus tr
43	116	11.7	589	14 AF184113	AF184113 Citrus tr
44	115.5	11.7	589	14 AF456773	AF456773 Citrus tr
45	115.5	11.7	672	14 AF501868	AF501868 Citrus tr

RESULT 1

ALIGNMENTS

AR138306
 LOCUS AR138306 597 bp DNA linear PAT 16-JUN-2001
 DEFINITION Sequence 14 from patent US 6197948.
 ACCESSION AR138306
 VERSION AR138306.1 GI:14479815
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE
 1 (bases 1 to 597)
 Authors Zhu, H.-Y., Ling, K.-S. and Gonsalves, D.
 Title Grapevine leafroll virus (type 2) proteins and their uses
 JOURNAL Patent: US 6197948-A 14 06-MAR-2001;
 FEATURES
 source
 1..597
 /organism="Unknown"
 /mol_type="unassigned DNA"

ORIGIN
 Alignment Scores:
 Pred. No.: 2,39e-97 Length: 597
 Score: 991.00 Matches: 198
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-09-613-486-15 (1-198) x AR138306 (1-597)

QY 1 MetGluLeuMetSerAspSerAsnLeuSerValIleThrAspAlaSerSerLeu 20
 DB 1 ATGGAGTGGATGTCGACAGCACTTACCACTGGATGATACCGACGCTTACGCTA 60
 QY 21 AsnGlyValAspIleValLeuSerValGluValGluValMetLeuValGluValGly 40
 DB 61 AATGTTGTCGACAGCACTTATCTGCTGATGATAAATGTTGGTCGAAAGGG 120
 QY 41 AlaProAsnGluGlyIleGluValValPheGlyLeuLeuLeuValAlaLeuAlaArg 60
 DB 121 GCTCTTACGAGGGATGATGAGAGTGGTGTGCTTACCTTACGACCTGCGGCAAGA 180
 QY 61 ThrThrSerProGlyValGluValAlaAspSerAspValIlePheSerAspSerPheGly 80
 DB 181 ACCAGGTCCTCAAGGTCGACGCGGCAATTCAGACGTTATATTTCAATAGATTGCGA 240
 QY 81 GluAcGAsnValValIleThrGluGlyAspLeuValValLeuAspGlyCysAlaPro 100
 DB 241 GAGAGGATGTCGTAGTAAAGAGGGTACCTTAAAGAGTAACTCGACGGGTGTGGCCT 300
 QY 101 LeuThrArgPheThrAsnLeuValGlyThrPheGlyValGlyThrPheThrGluAlaVal 120
 DB 301 CTCACCTAGGTCACCAATTAACCTTAAAGACGTTGGTGTGCTTACCTTACAGAGGCTTACGTT 360
 QY 121 AspPheCysIleAlaValIleValIleValIleValIleValIleValIleValIleVal 140
 DB 361 GACTTTGTTATCGCTTAAAGCAAAATTAACCCCAACCAACGCGCGCGGGAATTGGGG 420
 QY 141 IleProAlaGluAspSerValLeuAlaAspPheLeuGlyValIleValIleValIleVal 160
 DB 421 ATTCAGGTCGAGATTCGATTAAGTACCTGACGATTTTTCGGGTATCTTGCCTCAAGCTCTCT 480
 QY 161 GluLeuGluGlnIleSerArgIleValPheAlaSerMetValAlaLeuValIleValIleVal 180
 DB 481 GATTTACGACCAAGTAAAGAGATGTTCCGAGTATGTCGCTTAAAGAGAGGTTGGA 540
 QY 181 ValValAspThrProValSerAsnLeuValGluValGlyValGlyValGlyValMet 198
 DB 541 GTGTAAATATACACAGTGAAGCAATCTGCTCAGCTAGTAGAGAGGAGATTATG 594

RESULT 2
 AF039204 15000 bp RNA linear VRL-22-MAY-1998
 LOCUS AF039204
 DEFINITION Grapevine leafroll-associated virus 2 methyltransferase/helicase

ACCESSION AF039204
 VERSION AF039204.1
 KEYWORDS
 SOURCE Grapevine leafroll-associated virus 2
 ORGANISM Grapevine leafroll-associated virus 2
 REFERENCE
 1 (bases 1 to 15000)
 Authors Zhu, H.Y., Ling, K.S., Gosczyński, D.E., McPerson, J.R. and Gonsalves, D.
 TITLE Nucleotide sequence and genome organization of grapevine leafroll-associated virus-2 are similar to beet yellows virus, the closterovirus type member
 JOURNAL J. Gen. Virol. 79 (Pt 5), 1289-1298 (1998)
 MEDLINE 98264507
 PUBMED 9603345
 AUTHORS Zhu, H.Y., Ling, K.S. and Gonsalves, D.
 TITLE Direct Submission
 JOURNAL Submitted (18-DEC-1997) Plant Pathology, Cornell University, New York State Agricultural Experiment Station, Geneva, NY 14456, USA
 FEATURES
 source
 1..15000
 /organism="Grapevine leafroll-associated virus 2"
 /mol_type="genomic RNA"
 /db_xref="taxon:64003"
 <..7423
 /function="replication and unknown functions"
 /note="larger than 277 kDa; ORF1a; contains domains for a helicase; identified by sequence comparison"
 /codon_start=2
 /product="methyltransferase/helicase polyprotein"
 /protein_id="AAC40855.1"
 /db_xref="GI:3123910"
 /translation="ADVYAMLRVYCGKPLVLMRSVITPDGRCYLAHMYLCAFYCR
 PRESIDYALGMPTVAVRLACVKEKVFAGAGIDLRGYTSRNVHCYDASVAYYFR
 NLSGRIGSGSPSTLTSTVITKLSGLGCGPKXIAFAFCDLIRYVAVDSGGIOSS
 VKTKEDHRTVEERAGSVYQPOKSIDERGCGRVSGFSLVNLVNLVEGRKVA
 AGLLRFRYGDMDPFRSPSTOAGHLLVGRVSSRCEILSPSNFLRYVLPJCSGD
 YAMPSFPAAGKRFPLVITRITRYPGFCLYACACACAPLIRGPKRFDIAGPPTAA
 KLRNMSVELSGRSGLNLGAYTSRGVPHCDYKPTKDIRLMSAVTAGDQGEVY
 PSDITPANKQKTIENVYDLVYGTSLSLKSLIEKQIDFKDVOGLKORFIVKVPY
 MEATQNSLTRFYPQPELKFSSHSDDPAAASRLBNELVRLCGNSVSDIGCPPL
 FLHAKTORVAVCPVLVDGKDCARVVDQYSNVRLDGLDEGRNVDIICPPL
 GACHSSAMMMVQYVDASLYEICAMIKKKSRITVIMVTGPELDERECVMSLD
 CELEVDVADVVMYKFGSSCYSHKSLIKDINTPTYLIGGLFVEMVEYMGVYF
 KITKSEVSGISCTLLKLYRRANSVYVYIKPRPKRMLPGVDITVLSKPSVY
 FIVVNCNSAVSKTFEWMYSFSSKSRVITSGKTIHKDYNLDKTYESPAAVNLAS
 GVRSPLDSEYLAKNLSHPSGDSPTLEAFVREKIRNMTLNFNRLQIVARVAFAT
 LNVSPFLDSTLESITDFAECKVALELDELGLRAENENKIRNLNGLSIAKLSAI
 VVDIDSKPSPKQVGNSSSENADREVQPGIARGSGRNVGGLFHRVDSALRIKVA
 TPOORIKSVYRFLDSAVFLDYNQNLFFILFLESGSCMPAFIANKGDSLRVSA
 VCAAYEVATSCNASVAKAKWITFAAIVCAAMFVSGSGSGGREGYKSIHYTVGLF
 DITPEPDSYCLEIYLSNAICGATVTLSSGSSISLNFILQITGFSLEVVAVRVR
 VTHGSLTITDVIKGVFQIYSHLNTQNTGVAQSAFISGLVPLVIAKCVSLIFIL
 RTDTISGFIHGISFPLSLTLKFLKGLVDLSLSSITGGVDSKHKYKKTQTAIR
 TTVMOYFVAVNALSKIAAKIYNFTSRTPFKDSSSDGACEYFSCDBSGGLK
 GGSYGFSLIAFFSRIMGARRLIVKHKECGKLFEFLSLHSEFRFRVFGKNTDV
 GYVDFLPTGIVETLSIECDQIEELLGDGLGDADSLTDNAYEFESDFPLASIEEP
 PFAKRGKSKNTAIIALEYANLPRIVYSCSKSPDLFAFELASALIEKKEVFR
 KGVIAVRYTORFLFRSRMLGANNVSVSPDLPLMKRKYSGFSLASVAVPL
 RGSYWCVAERREKFRALVIGLSLTPVIVAKKCEELVSSARBYEIKIFLR
 QKTVSLNFTFCFLPSSDNDSSASAGLKGASRMTLPHLYVLAASALISGEGKTL
 LSHHNLFLCFLAVDQVNLIVLGLGSFVPIFSLFAMLLQDPRFVYEBKVTYA
 PEFFLKCSPPRALPLKGFPECVANSTVSTVRLRCFVKMLIKRGRGLDGRGLH
 ROKAVVIPSNRVVDGVERLSVKMGVVALTEELRIEDDSAVIEKLNRRRNDTN
 DDEFTRPAHEMOEVTTFCSKANSAGLIERVAVLEDADIKSKRLKTVEMVRSKSTT
 SEEVAVALSDDEAVEIEISVADBRDSEFKTVISEYLNRLNSSFEPFKPIVVDNDGTG

CDS

polyprotein gene, partial cds; and RNA-dependent RNA polymerase, putative transmembrane small hydrophobic protein, 65 kDa chaperone protein, 63 kDa protein, 25 kDa diverged coat protein, 22 kDa coat protein, 19 kDa protein, and 24 kDa protein genes, complete cds.
 AF039204.1 GI:3123909

CDS
 GLTNVAREFYVMOELAFELHSKLCITYDQRLVNFDRSVAPCSEDAOLYKRNKSTI
 VQKEVRLHIKDFHDFJEDKISINKRRRGVNLVHNDLAFVLSNLSLADYFPRS
 FVETNSVDILLYEAPPGGKTTTLIDSLFKVFKGEVSTMLITANKSSQVBLIKRVE
 KEVSNIECQKRKRKRSPKRSIYIDAYLMHNRGCDADVLFIDECFWHAGSVLACIIEF
 TRCHVMIFQDSQIOIYIERNELDKCYGLDFVDLOCRVGNISYCPMDYCAMLS
 TVYGNLJATYKGBEGKSSRMNREINSDVDVPGSTFLCMLOSEKLEISGPIKRG
 LITLANTITHEAGETAYARNIVRLKODDEPKSRHITVVALSRPTDSITVYVALAR
 RGDATDAIOKAEIVNKRKRVPTSGSGITINLVKVDENDSRCAASAPISVINDP
 LNEVPGTAVIADGDLSDPSTGPECGASIVVRUNISSNITDHDQRV"
 <7422..8801
 /function="replication"
 /note="RDRP; 52 kDa; similar to RNA polymerases of other
 closteroviruses; presumably expressed via +1 ribosomal
 frameshift"
 /codon_start=1
 /product="RNA-dependent RNA polymerase"
 /protein_id="AAC40856.1"
 /db_xref="GI:3123911"

CDS
 /translation="SVYRQAIIPRRKPSLOENLYSFEARNVSTCDRATSMGEA
 MAMNCLRCRCDLAPSLRDVLSITRSGIEOMLEKRTPSQIKALMKDYESPLEIDE
 ICFKLMTKRDADAYKLDSSCLTGHSAONIMFERKSNALIESPIFNEVNRIMCCLKP
 NIKFTEMTNRDFAVSVMNLGDDVYHICEVDPSKIDKSODAFVAFEBVYKELGV
 DEBLALVMCGERLSIANTLDQLSFTIENQKSGASNWIGNSLVTLLISLYDVR
 NFPAIYISGDSLIIFSRSEISYADICTDMGEFTEKPSGVPYFCSEKRVVWCGHFT
 FVPDPYKLFVKLGAVKEDVMDPLFEFTSFKDLTSDPNDERLIQRLALVALKVEVQ
 TGNTTALSVIHCIRSNFLSFKLYPRVKQVFTYSVKALKLSGCLSFDSFMTPEFG
 QAYWMDDE"
 8865..9035
 /note="6 kDa; probably membrane-associated; similar to
 small hydrophobic proteins of other closteroviruses"
 /codon_start=1
 /product="putative transmembrane small hydrophobic
 protein"
 /protein_id="AAC40857.1"
 /db_xref="GI:3123912"

CDS
 /translation="MNVQLQPECLFLNLAVAVTIFILLVRVILKSRQKHEAPV
 PVVRGGGFSTYV"
 9051..10850
 /note="p65; HSP70; similar to heat shock 70 proteins;
 identified by sequence comparison"
 /codon_start=1
 /product="65 kDa chaperone protein"
 /protein_id="AAC40858.1"
 /db_xref="GI:3123913"

CDS
 /translation="MNVFGLDPGTTFSTVCVYKDRYFSKONNSAYIPTLYLFSDS
 LNMFTGYEASLMSNLKYKGSFYDLKRWQCDSSNDAYDLRKPHYSVRLKIGSG
 LNETVSIKSGTGVSEAHLPGLIALFIKAVISCAEGAFACCTGIVCSVPARYDSQ
 NMTPDCCVSLSGVOCVVMNEPAAALISACNSANLAVDFGGGTDDVILISYR
 NMTFVPAASGDNLGSRDVPRAFLTLFSLTISPLDITDINIKESLSKMDATVY
 TLRGVDRKEDVRANKNLTLSVMLPYNRFLTKLESTLKYAASNAESARVKKDYLVI
 GGSSYLPLGLADVLTKQSVDRILKSDPRAVAVGCLYSGSGGLLILCAHAT
 VALIDRSCHOIICAPAGAPILPFGSGMPLVYARVAKQSQSRVAFEBGVKCPNRIKC
 GANIRFPDIIQVTDGSAVPTVFYDFSISSGAVSFVVRGEGQVSLTGPANFSYV
 ALGSRVRELIHINNKVFLGLLHRKADRIITKDEAIRVADSIDIADVLKEYSY
 AASALPDEDVEYELLKRSVKVLRSLREEIP"
 10777..12432
 /note="p63; putative heat shock protein 90 homolog"
 /codon_start=1
 /product="63 kDa protein"
 /protein_id="AAC40859.1"
 /db_xref="GI:3123914"

CDS
 /translation="MSNVSMESLFFKFFYGEADMKYLSRSIAHSEIKTLPIRLYG
 GRVYKSFESALNSPEOELGLFISERVGSKLGIIVESAAYLITPKRYKFA
 ETCSVDKGEQKYSMEDVNMFWLSTLNDNDKLTETQCSLNSGCELTINPDKRE
 VALTFKDDIADDTGAAVBCRVGDYLYVMSLFEGQTQSGSNNITLYKCYEYRT
 YLGGTDLFTFAPDRIPILITGLIDYFCEYVNFYSYRANVDFEFLIANWPLISDV
 VFQWKPAPVRLLEFELSAELLLEVPLLIDSQVAVGHILRVESYTSDDPAILDAE
 DKLBAIKSNPRLSTAQWGFQYCEGFTAQSRVQSGVYGLTSDYSGGFEINIK
 DYKEFPDLGRELPNVSLRQFNAGARAEAKFIKPNINISRPISRLNVRREYINIK
 DVEFHANRSLTEBELLINNIISVDVRLCAERACNTLPSAKRPSXNHKSNISSRQE
 RRIQDPLVYKDTLYEFQHKRAGWGSSTRDLSRADHAKGSG"
 12344..13015
 /note="p25; coat protein duplicate"

CDS
 /codon_start=1
 /product="25 kDa diverged coat protein"
 /protein_id="AAC40860.1"
 /db_xref="GI:3123915"
 /translation="MSNTSVYPVGGLEBALETSGVVLITRKAVDKFENELKENYSSV
 DSSLSDSEVVELEKSESFKSELSTDEHFPVHIFFLIRCAKISTSEKVKVGRH
 TYVVDGKYTVLDAWVFNWMSLTKYKRVNGLRAPFCACEDLYLVAIPMSRFKTK
 AVKMGGLPVGKEYIGADPLFGSTKMSDHDRAVSIVAARNAVRSATGGERKIVSLY
 DLGRY"
 13084..13680
 /note="p22"
 /codon_start=1
 /evidence=experimental
 /product="22 kDa coat protein"
 /protein_id="AAC40861.1"
 /db_xref="GI:3123916"

CDS
 /translation="MELMSDSNLSNLYTTDASSINGVDYDKLISAEVKMLVOKAPNE
 GIEVFGLLIYALARTSPKVRQADSDVIFSNRFGRRNVVTEGDLKIDGCAPIIT
 RPTNKLRTFRTFTEAVVDPCIAVKKHLPOLNAAALGIPAEBSYLAADFLGTCPLKS
 ELQOSRKMFAAMVAKTEGGVNTVPENLRQIGREVM"
 13680..14165
 /note="p19"
 /codon_start=1
 /product="19 kDa protein"
 /protein_id="AAC40862.1"
 /db_xref="GI:3123917"

CDS
 /translation="MEDYEKSESLILRTNLTMLVYKSDASVELPKLIGYLRV
 SGRGVTCNREELTRDPEGNHHTVIRSLIOYSESAFEBEENNSDCVKKFLETGSV
 FMPFLRSTTKGRAVRLHRTFPEANNFPGSHGEMEXCLQVLTETSIIDSFCBERN
 R"
 14167..14784
 /note="p24"
 /codon_start=1
 /product="24 kDa protein"
 /protein_id="AAC40863.1"
 /db_xref="GI:3123918"

CDS
 /translation="MNVIVSPEBAEDILKSTDMLRNIDSGVLSKECIKAFSTTRD
 LHCAKASIQMGVDTGLYORNCARGLIDVSNLRLOPVREVAVAFQCDKPEKELV
 AFTIRKYVELTGVGREAVERKRSRLTKVLRKSLMAFYMSPRAKNAEMLELKS
 PVKIFRDLILDEVETNELCAEDDVHDKVKNNGENDHLELQDC"
 14785..15000
 /note="similar to 3'UTRs of other monopartite
 closteroviruses"

ORIGIN

Alignment Scores:
 Pred. No.: 1,696-95 Length: 15000
 Score: 991.00 Matches: 198
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 Gaps: 0
 DB: 14

US-09-613-486-15 (1-198) x AF039204 (1-15000)

QY	1	MetGluLeuMetSerAspSerAsnLeuSerAsnLeuValIleThrAspAlaSerSerLeu	20
DB	13084	ATGAGATTGATGTCGACAGCAACTTAGCAACCTGGTGATACCGAGCGCTCTAGTCTA	13143
QY	21	AenGlyValAspLysLysLeuLeuSerAlaGluValIGluVSMetLeuValGlnLysGly	40
DB	13144	AATGGTGTGACAGAGAGCTTTATCTGCTGAAGTTGAAAATGTGGTGAAGAAAGGG	13203
QY	41	AlaProAsnGluGlyIleGluValValPheGlyLeuLeuLeuTyraIleuValaIlaArg	60
DB	13204	GCTCCCTAAGAGAGGATTAAGAAATGCTTCCTACTCTTACGACTCGCGGCAAGA	13263
QY	61	ThrThrSerProLysValGlnArgAlaAspSerAspValIlePheSerAsnSerPheGly	80
DB	13264	ACCACGCTCTCTTAAGATTGAGCGCCAGATTGAGAGTTATTTTCAATAGTTTCGGA	13323
QY	81	GluArgAsnValValThrGluGlyAspLeuLysValLeuAspGlyCysAlaPro	100

Db 13324 GAGAGGAATGTGTAGTAACAGAGGCTGACCTTAAGAAAGTACTCCAGCGGTGTGGCCT 13383
 QY 101 LeuThrArgPheThrAsnLysLeuArgThrPheGlyValArgThrPheThrGluAlaIleVal 120
 Db 13384 CTCACGTAGTTCACCTAATAAAGTAAAGCTTGGTGTGCTACTTCACTGAGGCTTACGTT 13443
 QY 121 AspheCysIleAlaIleTyrLysHisLysLeuProGlnLeuAsnAlaAlaGluLeuGly 140
 Db 13444 GACCTTTGTATCGCGTATAGCACAATATACCCCACTCAACGCCGCGGGGGAATTGAGG 13503
 QY 141 IleProAlaGluAspSerTyrLeuAlaAlaAspPheLeuGlyThrCysProLysLeuSer 160
 Db 13504 ATTCACGTGAAGATTGCTACTTACCTGACAGATTCTTGGGTACTTCCCGAAGCTCTCT 13563
 QY 161 GluLeuGlnGlnSerArgLysMetPheAlaSerMetTyrAlaLeuLysThrGluGlyGly 180
 Db 13564 GAATTACAGCAAAAGTAGAGAGATGTCGCGAGTATGACCTCTAAAAACTGAAGGTGGA 13623
 QY 181 ValValaAsnThrProValSerAsnLeuArgGlnLeuGlyArgArgGluValMet 198
 Db 13624 GTGGTAATATCACCACTAGACATCTGCTACGCTAGGTAGGAAGGAAATTATG 13677
 RESULT 3
 LOCUS AR138299 15500 bp DNA linear PAT 16-JUN-2001
 DEFINITION Sequence 1 from patent US 6197948.
 ACCESSION AR138299
 VERSION AR138299.1 GI:14479808
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 15500)
 AUTHORS Zhu,H.-Y., Ling,K.-S. and Gonзалes,D.
 TITLE Grapevine leafroll virus 2 proteins and their uses
 JOURNAL Patent: US 6197948-A 1 06-MAR-2001,
 FEATURES
 source 1..15500
 /organism="Unknown"
 /mol_type="unassigned DNA"

ORIGIN
 Alignment Scores: Length: 15500
 Pred. No.: 1.76e-95
 Score: 991.00
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 100.00%
 Gaps: 0
 DB: 6

US-09-613-486-15 (1-198) X AR138299 (1-15500)

QY 1 MetGluLeuMetSerAspSerAsnLeuSerAsnLeuValIleThrAspAlaSerSerLeu 20
 Db 13584 ATGAGGTGATGTCCGACAGCAACCTTAGCAACCTGGATGATACCGACGCTCTTAGTCTA 13643
 QY 21 AsnGlyValAspLysLysLeuLeuSerAlaGluValGluLysMetLeuValGlnLysGly 40
 Db 13644 AATGCTGTGCAAGAAAGCTTTATCTCTGAGAGTTGTAAGAAAGTTGGTGCAGAAAGGG 13703
 QY 41 AlaProAsnGluGlyIleGluValValPheGlyLeuLeuLeuTyrAlaLeuAlaAlaArg 60
 Db 13704 GCTCTTAACGAGGATATGAAAGTGTGCTGCTACTCTTACGACACTGCGGCAAGA 13763
 QY 61 ThrThrSerProLysValGlnArgAlaAspSerAspValIlePheSerAsnSerPheGly 80
 Db 13764 ACCACGTCTCCCTAAGGTTCAAGCGCAGATTCAGACGTTATATTTTCAAAATAGTTTCGA 13823
 QY 81 GluArgAsnValValValThrGluGlyAspLeuLysValLeuAspGlyCysAlaBro 100
 Db 13824 GAGAGGAATGTGTAGTAACAGAGGATGACCTTAAGAAAGTACTCGACGCGGTGTGGCCT 13883
 QY 101 LeuThrArgPheThrAsnLysLeuArgThrPheGlyValArgThrPheThrGluAlaIleVal 120

Db 13884 CTCACGTAGTTCACCTAATAAAGTAAAGCTTGGTGTGCTACTTCACTGAGGCTTACGTT 13943
 QY 121 AspheCysIleAlaIleTyrLysHisLysLeuProGlnLeuAsnAlaAlaGluLeuGly 140
 Db 13944 GACCTTTGTATCGCGTATAGCACAATATACCCCACTCAACGCCGCGGGGGAATTGAGG 14003
 QY 141 IleProAlaGluAspSerTyrLeuAlaAlaAspPheLeuGlyThrCysProLysLeuSer 160
 Db 14004 ATTCACGTGAAGATTGCTACTTACCTGACAGATTCTTGGGTACTTCCCGAAGCTCTCT 14063
 QY 161 GluLeuGlnGlnSerArgLysMetPheAlaSerMetTyrAlaLeuLysThrGluGlyGly 180
 Db 14064 GAATTACAGCAAAAGTAGAGAGATGTCGCGAGTATGACCTCTAAAAACTGAAGGTGGA 14123
 QY 181 ValValaAsnThrProValSerAsnLeuArgGlnLeuGlyArgArgGluValMet 198
 Db 14124 GTGGTAATATCACCACTAGACATCTGCTACGCTAGGTAGGAAGGAAATTATG 14177
 RESULT 4
 LOCUS GLAV4131 8590 bp RNA linear VRL 08-JAN-1998
 DEFINITION Grapevine leafroll-associated virus 2 genes encoding RNA polymerase and coat protein, hep70, hep90 gene and ORF2, ORF7 and ORF8.
 ACCESSION Y14131
 VERSION Y14131.1 GI:2765391
 KEYWORDS coat protein; heat shock protein 70-like; ORF1; ORF2; ORF3; ORF4; ORF5; ORF6; ORF7; ORF8; RNA polymerase.
 SOURCE Grapevine leafroll-associated virus 2
 ORGANISM Grapevine leafroll-associated virus 2
 Closes: sRNA positive-strand viruses, no DNA stage; Closteroviridae; Closterovirus.

REFERENCE 1
 AUTHORS Abou-Ghanem,N.
 TITLE The nucleotide sequence of the 3' terminal region of grapevine leafroll associated closterovirus 2
 JOURNAL Unpublished

REFERENCE 2
 AUTHORS Abou-Ghanem,N.
 TITLE Direct Submission
 SUBMITTED (01-JUN-1997) N. Abou-Ghanem, Universitat degli Studi di Bari, Dipartimento di Protezione delle piante e Centro Studio Virus e virosi delle colture mediterranee, via Amendola 165/A, I 70126 Bari, ITALY
 REVISED BY (3)
 3 (bases 1 to 8590)
 Abou-Ghanem,N.
 DIRECT SUBMISSION
 SUBMITTED (07-JAN-1998) N. Abou-Ghanem, Universitat degli Studi di Bari, Dipartimento di Protezione delle piante e Centro Studio Virus e virosi delle colture mediterranee, via Amendola 165/A, I 70126 Bari, ITALY
 On Jan 10, 1998 this sequence version replaced gi:2369864.

COMMENT
 FEATURES
 source 1..8590
 Location/Qualifiers
 1..8590
 /organism="Grapevine leafroll-associated virus 2"
 /mol_type="genomic RNA"
 /db_xref="taxon:64003"
 /lab_host="N. benthamiana"
 1165..2427
 /note="ORF1"
 /codon_start=1
 /product="RNA polymerase"
 /protein_id="CAA74561.1"
 /db_xref="GI:2369865"
 /db_xref="GOA:O39852"
 /db_xref="SPTREMBL:O39852"
 /translation="MNGEAMANNCRPRCDLAFSLINDVTSIRSGIEQWLGKRP
 SRKALKMDVBSPLIDDETCRFLMVKRDAVKDLSCTTHGSAONIMFRKSLNA
 IISPLFNEVKRIMCCLKPNIKRFTENRDRFASVSNMGDDYTHISEVDPSRYDK
 SDAFYKAEVEYMKELGVDEBLAIIMKGEELSIANTLDGSLFTTEORSGASNT
 WIGNSLVITGIISLYVDVNFALYISGDSLIIFRSEISNVAADICTDMGEITFMS
 PSVPYFCSKFFVMCGHRTFFVLDPYKLFLKGAIVEDVSMDSLFETFTFSFKDLTSDFN

[illegible]

```

CDS
      /db_xref="SPREMBL:O39857"
      /translation="MEIMSDSNLSNVIYTPASSINGVDKLLSAEAVTKMLVKGAPNE
      GIVAVGLIYLAALATTSKVRADSDV FNSFGERNVYVEGDLKTVLDGCAPLT
      RFTNKLRTFGRTTEAVYDRCIAKYKRLPOLNAAELGIPAEDSYLAADPLGTCPKLS
      ELQSRMRFPASMYALKTGEGVAVT PVSNLQLGRREV"
      7306..7791
      /note="ORF7"
      /codon_start=1
      /product="hypothetical protein"
      /protein_id="CAA74567.1"
      /db_xref="GI:2369871"
      /db_xref="SPREMBL:O39858"
      /translation="MEDYEKSSLLILRTNLMTMLVKSDAVLEPLICGLYLRV
      SKRGVYCCRRRELTTRPESGHHYVTSRILOVDSSEAFEPNNSDCVVFKEFTGSGS
      FWFPLRSTKGRVAVRHILTFEAVNFFFGHCGIMEYCLQVLTETESILDSPEERN
      R"
      7793..8410
      /note="ORF8"
      /codon_start=1
      /product="hypothetical protein"
      /protein_id="CAA74568.1"
      /db_xref="GI:2369872"
      /db_xref="SPREMBL:O39859"
      /translation="MRVIVAPYEADILIKRSTDMLRNIDSGVLSSTKEICAKFTITRR
      LHCKASYQMGVDYGLYQRCNCAEHLIDVPSNIRLQPIVREKVAIFGCKDEPELIL
      AFITRRKVELTGVGMREAVVRKMSLTKTIVNKSLEMAFYMSRPMAGNAWMLKRF
      PVKIFRDLLIDVETINELCAEDDVHVVKVKNENEDNDLELQDEC"

```

[illegible]

US-09-613-486-15 (1-198) X GLAV4131 (1-8590)

QY	1	MetGluLeuMetSerAspSerAsnLeuSerAsnLeuValIleThrAspAlaSerSerLeu	20
Db	6710	ATGGAGTTGAGTCCGACAGCAACCTTAGCAACCTGGATGATPACCGAGCCCTTAGCTCTA	6766
QY	21	AsnGlyValAlaAspLysLysLeuLeuSerAlaGluValGluLysMetLeuValGlnLysGly	40
Db	6770	AATGGTGTGCACAAAGAACCTTTTATCTCTGCTGAAGTTGTAAAAAATGTGTGGTCAGAAAGGG	6823
QY	41	AlaProAsnGluGlyIleGluValValPheGlyLeuLeuLeuIleValAlaLeuAlaIleArg	60
Db	6830	GCTCTTAACGAGGGATATGAAAGTGGTGTCCGTCTACTCTTACCTTAGCGACTCCGGCGAAGA	6888
QY	61	ThrThrSerProLysValGlnArgAlaAspSerAspValIlePheSerPheSerPheGly	80
Db	6890	ACCACGTCTCTTAAGGTCCAGCGCCGCGAGTTACAGCGCTTAATATTTTCAAAATAGTTTCGA	6944
QY	81	GluArgAsnValValValThrGluGlyAspLeuLysLysValLeuAspGlyCysAlaPro	100
Db	6950	CAGAGAGATGTGGTAGTACAGAGGGGTACCTTAAAGAGTACTCGACGGGTGTGGCCT	7004
QY	101	LeuThrArgPheThrAsnLysLeuArgThrPheGlyArgThrPheThrGluAlaIleVal	120
Db	7010	CTCACTAGGTTACACTAATPAACTTAGAAGTTCGGTTCGACTTCACTGAGGCTTAGCTT	7066
QY	121	AspPheCysIleValAlaTyrLysIleLysLeuProGlnLeuAsnAlaAlaGluLeuGly	140
Db	7070	GACITTTGTATGCGCTATATAGACAAATTAACCCCACTCAACCGCGCGCGGAATTTGGGG	7122
QY	141	IleProAlaGluAspSerTyrLeuAlaAlaAspPheLeuGlyThrCysProLysLeuSer	160
Db	7130	ATTCCAGCTGAAGATTCGTACTTACTACTGCAGATTTTTCGGTACTTGGCCGGAAGCTCT	7188
QY	161	GluLeuGlnGlnSerArgLysMetPheAlaSerMetTyrAlaLeuLysThrGluGlyGly	180

Db 7190 GAATACAGCAAGTAGAAGAAATGTCGACGATGATGACGCTCTAATAAACTGAAGGTGGA 7249

Qy 181 ValValAsnThrProValSerAsnLeuArgGlnLeuGlyValArgGlnValMet 198

Db 7250 GTGGTAATACACCGAGGAGCAATTCGGCTACACTAGGTAGGAGGAAAGTTATG 7303

RESULT 5

AF314061

LOCUS

DEFINITION Grapevine rootstock stem lesion associated virus

mechylitransterase/helicase polyprotein, RNA-dependent RNA

polymerase, p6, HSP70-like protein, HSP90-like protein, coat

protein duplicate, coat protein, p19, and p24 genes, complete cde.

AF314061

AF314061.1 GI:24636914

16527 bp ss-RNA linear VRL 06-NOV-2002

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE 1 Grapevine rootstock stem lesion associated virus

2 Grapevine rootstock stem lesion associated virus

3 Viruses; ssRNA positive-strand viruses, no DNA stage;

4 Closteroviridae; Closterovirus.

1 (bases 1 to 16527)

2 (bases 1 to 16527)

3 (bases 1 to 16527)

4 Nucleotide sequence of grapevine rootstock stem lesion associated virus

Unpublished

2 (bases 1 to 16527)

Zhang, Y. and Rowhani, A.

Direct Submission

Submitted (16-OCT-2000) Plant Pathology, University of California, One Shields Ave., Davis, CA 95616, USA

Location/Qualifiers

1. 16527

/organism="Grapevine rootstock stem lesion associated virus"

/mol_type="genomic RNA"

/db_xref="taxon:167634"

107. 8902

/note="327 KDa"

/codon_start=1

/product="methyltransferase/helicase polyprotein"

/protein_id="AA063466.1"

/db_xref="GI:24636915"

/translation="MSSLAISALPSAAVAGQSFGQPAATVAKSFYMTSLPSLQCTTCQ

EFSSLFICLASFQKIKITFLSFLSVREIALFSELTIGFCCEFRILGVGHPFFVAF

OCGLSVCLGCVNGIPALVNISSPFLDVALSLMVGVVTPVTKVLSKRAKLNKRA

SILABAARIRSKVGNKEVRSIVQPLPAGCLRFSONKNGTELIIGRDVAVARVT

AAPIVRRGGCGRVPLIPLVAVSPGVAVRCVACGAIACFKKMLASVAVANDSRAE

KLUSRRRKLQOQNFNSRVAKKADVTASRDVAEGVTPPHHIFGSPSPSPSPSP

SGMASSVIREPAIVTCGASAPAIKEVPRVSKADSPAPAPVREKLTIRVAKRGA

IVISGVVAEVLNKIREGLRVFGDMKFRFSTHGCKIVANVRNTSVMLNCTIV

NGEKSVAI PCSVDVYAMLRVSGGKPLVMSRECPDGLCVANHRVYICAFQCP

RESDPFLGSPSVALKALVLRFGEDALNIGVGYVSRAPFCDSDSKARSLML

AGYVGDEVIHSTLASVTLKRGESADRPKISIFGSPCELRVYESVNDAPLRRE

KYLRYVERNRGRTSADVGSNVRKRDRIIDREKSHLAGSVNKKKRIAGVLRYS

VGGDDFNHSPFLQGHHLVLRKRTSSVCCELSLTPDRILRRVITPCSDYAMSR

LAAGRFPLVMTREDYKNGCYLAACRYAAVLLKGRPAVFDIGANPTAKSRMY

SVLGRSLNLNIGSFTSRGI PHCDYAAVAVYDRLFMSAIVAGEKEGEVAASTI

PKAKAHEKSEKSEKSEKSHAKGRGAGVLEKFLDPVLSASRLFSVCADQRIK

SVYALIDTAVSFLLDONENYLSFIMRVLSKGYSCMFSYLANGSLANRASAQIVREI

AYLCANVSYSKAKVVTTPSAACAMNENSCGSGDREYKSYMARVYQVLETFEED

SSYDPLELSAICCAITVLRSSGYSVSNALLOITTSFELSLIVRVAVVTHSLST

TATDQMRVSFQIISHLVNTGNALVQIAPLITVAVLLYKCKGGLFMSGDTTSG

FLHGVSESPFTNLILYRGLVDELAAVGVGVDSNKLHPRDVTQSAIRISTVQIIP

TAVDGLKCAASKVSGIVGKKTNGEESRGDGDHEEYFSCDESETPGLRGDSRG

CDS

10373. 10543

/note="6 KDa"

/codon_start=1

/product="p6"

/protein_id="AA063468.1"

/db_xref="GI:24636917"

/translation="MDVQLQFCELYLNAQVAFTVFILLIRVYKSRFNVHETPI

SAVRQGSFTV"

10552. 12351

/note="65 KDa"

/codon_start=1

/product="HSP70-like protein"

/protein_id="AA063469.1"

/db_xref="GI:24636918"

/translation="MVFGLDRCGTFTSTVCYVKGQVYSPKONNSAYIPTLYLSEET

NKMTGVEEASLHNRKARGSPRLKRWGDSNPFSDYDRLLKPHVAVPYKIGTG

LNDFVITIGVGTIVSEVHLPEGLISFKIAMSCEANAFCCTCGVICSVPANDVSG

RNFPTQCVLSGYQCVVWNEPSAALACNLINKSNANLAYIDFGGTFPVISISYR

NNTFVVRASGGDLNGRDVDRTPFLTHLFSLSLEPDLSDLSINKESLSKDAIV

TLKSGDDKEDVAVKSLITVLMVYVRNTLILATIKSYAKNGGARSVCDAIVL

IGSSSYLPGLEDILSKHRDRILIKVADPAAVAVGCAIYSGLSGOGLIWDCAHT

IALAGSCDQIICAPAGAPIPSGTPPLVPAANSORQIAPFGEVGYCGRNKKIC

GSNIRKLDGLGHTSVASVPTPTVLDITISVAGISVWNGPEKEXVLTGTPANSSV

VJGTRVRLGHTSLSNKRALGLILRRKADRLRLPBEGEAVRYEIVAVADVLEKRSY

NASSLPDDVDVPELNGKSVQVLRGRGRLEPL"

12278. 13882

/note="62 KDa"

/codon_start=1

/product="HSP90-like protein"

/protein_id="AA063470.1"

/db_xref="GI:24636919"

/translation="MSNHSWGLSFKKYTSSEADWRKLSYTIANTNTLPDVOY

GRVYKSESA="PSRFEQELGLMLSEEDVGMCRGCTIVETANTQINPAPYPIA

EKCDPVKTEGQKTYMEDIWPMRNSNDINDIKILAEQGSJNSGCLINPDBBERF

VALTFKNELVDSNANIDCRVGLVLYANLFCRQSGKSGAGNISLVEKCAVIRT

YIENYDLYFSPRIPLLTGLIYDFCKEYNIYSTYKRVNDVRFELTNVMPILSIF

VROMVPAIDVRLPDLISAELTLEVPTSLVDSQVYGVHVARFVESYADPAIDALE

EKLDAIINKSNPILSLNQLMWGFCFYGGFRAAGRVVGRPVYATPNSVGGEINMK

PSILPFPSRLVRARILIGRIKRCRSKIPESISLKHHEFTLTKGSDVGVVEFL

SEVDADLSSIECDTEINLIGDFRMAWDSITDMWYFDPSPFLASVNGVLCAGR

GGRRVGMALILNVNYSILIMARAASKNSTFTSLIPSGLVDRIRNIPSNKESVVE

LVRKYVNGFLKSNRGLGLNREHYVNSFGELVPLLNKLSSETFIYRAVMSWCMV

CLASEREKEFALVLTGLGFYARLFGVGVCDKLSLRAVADRYVLEKRCDA

MNGIRCLPHPDNGDSTASGKGGSGOTFMGLLRVLRALSGVAGCCVSHHDL

FLCLSVLEBYEMIAKTIIGLSPSPRSALAPALLOPERFVSGKSTPLFLFLK

ISPRVPLIKGFEEVAVASPLLRGVRPLIAVKNLNRKGLRRNGRSGORAPV

VIPSNRVYNDGCEKLSKQVGEALREHLLESDLAVERKLNKRNRRNADSSPBR

PMBSRTDVYSPCKANSVGTALERAVLEDAKSEKISMAWDLVRKGVSDBEITN

LTSDDEALIEEYVSEERDSDPKTVRVEVILNRLNSPEPDIYVNDKSTGLTNAV

REFYMOGALFEIHSKLCAYDQLRIVNEERSAPCNEDAOIYVRNGSTIYQGVQ

RSNKLDPHDHDFEDGKISVRRRSNGVYHDLISPLANSFLAGYFERSFVNS

SVDLIYAPRGCKTTLTIDSPLKVRKEVSMILITAKSGQVEITLKYVEDISNI

EQQRKDRTRTKSITITIDALMHRCDDADIPIDCEVYHAGSVIACIEPFRCHV

MIPEDSRQIHYIERNEIDKCLYGDLDPEVQOCQVYNNVSRVCMVDCALSVYNDM

IATYGESEGKSNRINEINSVDLVPMDSITYCMLQSELEISKFIKRGSKLV

LTVHAGQETVAVNLVRLKFOEDPEPKSIRHITVALSRITSLTYNLVLAARGDATIC

DAIOKANELVNRVFPFSGSVINIDVKKVDYDNRCSAASPLSIVINDFLSDVNP

GTVVIDFGDLSADPSTGFBEGAGIYVRDNISSTITDDKRV"

8901. 10280

/note="53 KDa; translated via ribosomal frameshift"

/codon_start=1

/product="RNA-dependent RNA polymerase"

/protein_id="AA063467.1"

/db_xref="GI:24636916"

/translation="SAVRSOALPRRKPELOENTLYSEARNYNFSTCDRFTSAYFEGA

MANNCIRCFPLETFESSRLDVTISITSGIEIOMLEKRPISQIKALMIDIEPLIDE

ICRFKLTKRPAKTKLDSCTKHSAPNIMFHRKSNALSPFENKVRIMSCLP

NIKFTEMTNRDPAVSANMLGDDVYHIGVDFGSKYDSDAKVPEEMVYELV

DEELIATWGCERLISANTLOGQLSFTLENRKSQASVTMIGSLVTLGILSIYDR

HEEALYVSGDLSLIFSRNKISNYADIDADMGEFTKENSVPYFYCSALVAVGCHKIF

FVPDPYKFLVGLKAVKEDVSMDFLEETTSKDTLSPENDRVYQKAEALVAVYQ

SGNTALALSVIHICLRNPLSPSKLYPRVKQVYTSVKALFPGSGSLYSVYTPFG

HAWYMWDE"

10373. 10543

/note="6 KDa"

/codon_start=1

/product="p6"

/protein_id="AA063468.1"

/db_xref="GI:24636917"

/translation="MDVQLQFCELYLNAQVAFTVFILLIRVYKSRFNVHETPI

SAVRQGSFTV"

10552. 12351

/note="65 KDa"

/codon_start=1

/product="HSP70-like protein"

/protein_id="AA063469.1"

/db_xref="GI:24636918"

/translation="MVFGLDRCGTFTSTVCYVKGQVYSPKONNSAYIPTLYLSEET

NKMTGVEEASLHNRKARGSPRLKRWGDSNPFSDYDRLLKPHVAVPYKIGTG

LNDFVITIGVGTIVSEVHLPEGLISFKIAMSCEANAFCCTCGVICSVPANDVSG

RNFPTQCVLSGYQCVVWNEPSAALACNLINKSNANLAYIDFGGTFPVISISYR

NNTFVVRASGGDLNGRDVDRTPFLTHLFSLSLEPDLSDLSINKESLSKDAIV

TLKSGDDKEDVAVKSLITVLMVYVRNTLILATIKSYAKNGGARSVCDAIVL

IGSSSYLPGLEDILSKHRDRILIKVADPAAVAVGCAIYSGLSGOGLIWDCAHT

IALAGSCDQIICAPAGAPIPSGTPPLVPAANSORQIAPFGEVGYCGRNKKIC

GSNIRKLDGLGHTSVASVPTPTVLDITISVAGISVWNGPEKEXVLTGTPANSSV

VJGTRVRLGHTSLSNKRALGLILRRKADRLRLPBEGEAVRYEIVAVADVLEKRSY

NASSLPDDVDVPELNGKSVQVLRGRGRLEPL"

12278. 13882

/note="62 KDa"

/codon_start=1

/product="HSP90-like protein"

/protein_id="AA063470.1"

/db_xref="GI:24636919"

/translation="MSNHSWGLSFKKYTSSEADWRKLSYTIANTNTLPDVOY

GRVYKSESA="PSRFEQELGLMLSEEDVGMCRGCTIVETANTQINPAPYPIA

EKCDPVKTEGQKTYMEDIWPMRNSNDINDIKILAEQGSJNSGCLINPDBBERF

VALTFKNELVDSNANIDCRVGLVLYANLFCRQSGKSGAGNISLVEKCAVIRT

YIENYDLYFSPRIPLLTGLIYDFCKEYNIYSTYKRVNDVRFELTNVMPILSIF

VROMVPAIDVRLPDLISAELTLEVPTSLVDSQVYGVHVARFVESYADPAIDALE

EKLDAIINKSNPILSLNQLMWGFCFYGGFRAAGRVVGRPVYATPNSVGGEINMK

CDS
 NVEEFPDKIQREVPVNSLRQPNRGARHAEAFVFKKNIISKPISRLNIPPEFMYLNV
 DYFRHANSGLSEEEVLLINNISVDVRLCAERACSTLPSARFSGKHSSVPSLRQ
 KNVRDPLALRNSLYEFHGNRRV"
 13845..14516
 /note="25 kDa"
 /codon_start=1
 /product="coat protein duplicate"
 /protein_id="AA063471.1"
 /db_xref="GI:24636920"

CDS
 /translation="MSSGTGIAVGSSEALETSQVLSRKEAVDFKFNELNENYSSV
 DSRSLDQVEKLEKESFKSELSAETDEHVAHIFELLISALISIDSKYKYS
 KYVSEKTYTQDWWYFMVIGLTKQYKRVNGLRAFCACCELYLTVAPASERKTK
 AIGMKLPYKREYLGADFLISGTSKLSMDRRAVSIYAAKNVDRSAFNGERKIVSLV
 DLGRN"
 14604..15200
 /note="22 kDa"
 /codon_start=1
 /product="coat protein"
 /protein_id="AA063472.1"
 /db_xref="GI:24636921"

CDS
 /translation="MELMSDDNLSCIVITDASLNGVDKLLSAVITKMLYOKAPSQ
 GLETFGLLYALAAARTSPKVRADSDITFQNTIGDKIVTDEGLDKVLEBGCPLN
 RFTNKLRTGRFTBAVYDFCVAYKHKMFQNLAAALGIIPADSYLAADFLGACPKLS
 ELQOSRKMPASVYALTEGGVNTVPVNLRLGRREV"
 15200..15685
 /note="19 kDa"
 /codon_start=1
 /product="p19"
 /protein_id="AA063473.1"
 /db_xref="GI:24636922"

CDS
 /translation="MEICQNPBALVLTNTONTLLVKSDEVDNLEPKLIGYLYR
 SRGDTVGNREDVIRDFGAHHTVRSSTVYDESAYKEVNADCVKFLFETGDV
 FMPFIQSDIKGRARHLRTFFEANNFQSGHGTMEYCLKQVILTESVIESFCERN
 R"
 15687..16304
 /note="24 kDa"
 /codon_start=1
 /product="p24"
 /protein_id="AA063474.1"
 /db_xref="GI:24636923"

CDS
 /translation="MRVIVPYEAEDILRSTREMLRNIDSGLADTRECIRKASTLRD
 LKAKSYOMCUDTGCDDCAKRLIDVENSILRIAHPLIRKXATHPCHQDREIY
 AFTIRKIVLTGVAAREAVKIERTKISKVILNKLSLNAFFYSPAPAKRAEMLDLRYT
 PLKIRYDLDDVLTALCEDEIHVVVTKODENHDELQEDC"

ORIGIN

Alignment Scores:

Prod. No.: 6 44e-88 Length: 16527
 Score: 921.00 Matches: 180
 Percent Similarity: 95.96% Conservative: 10
 Best Local Similarity: 90.91% Mismatches: 8
 Query Match: 92.94% Indels: 0
 Gaps: 14

US-09-613-486-15 (1-198) x AF314061 (1-16527)

QY 1 MetGluUeuWetSerApsSerAenUeuSerAenUeuValIleThrAspAlaSerSerLeu 20
 Db 14604 ATGGAGTTGATGCCGATGATTAACCTTGAGCGGCTCGCTCAATACCGACGCTTTAGATTG 14663
 QY 21 AsnGlyValAspLysLysLeuLysSerAlaGluValGluLysMetLeuValGlnLysGly 40
 Db 14664 AATGGTCTCATTAAGAAATGCTGCTCGCGAGGCTCAAAAATGCTTGCGAGAAAGGGG 14723
 QY 41 AlaProAsnGluGlyIleGluValValPheGlyLysLeuLeuTyrAlaLeuAlaAlaArg 60
 Db 14724 GCGCCCGCCCAAGGATATTGAACAGTCTTTGGGCTACTGCTTACCGCTTCGACAGAGAA 14783
 QY 61 ThrTherProLysValGlnArgAlaAspSerAspValIlePheSerAenSerPheGly 80
 Db 14784 ACCACATCGCCCAAGTTCAGAGGGCTGATTCTGATTAATCTTTCAAAACACTTACGGT 14843
 QY 81 GluArgAsnValValValThrGluGlyAspLeuLysValValLeuAspGlyCysAlaPro 100

Db 14844 GACACAGCTGTGTGTCGACAGAAAGGTGATCTCAAGAAAGCTTTGAAGGTCGACACT 14903
 QY 101 LeuThrArgPheThrAsnLysLeuArgThrPheGlyArgThrPheThrGluAlaTyrVal 120
 Db 14904 CTTATATGATTCACTAACCAAGCTGCGAACTTTCGCGCGACGTTTACGAGGCGGTATGTT 14963
 QY 121 AspPheCysIleLeuAlaTyrLysLysLysLeuProGluLeuAsnAlaAlaGluLeuGly 140
 Db 14964 GATTTCGCGGTGTTCAAGCAATAGATCCCTGAGCTCAACCGCGCGCAATTCGGG 15023
 QY 141 IleProAlaGluAspSerTyrLeuValAlaAspPheLeuGlyThrCysProLysLeuSer 160
 Db 15024 ATACACAGAAAGTTGTCCTTACGCTGCGGACCTTTTAAAGTCTTGCCGCAACTCTC 15083
 QY 161 GluLeuGlnGlnSerArgLysMetPheAlaSerMetCysTyrAlaLeuLysThrGluGly 180
 Db 15084 GAATTACACAAAGTATGAAATAAGTTTGCAAGATGATGACCTTTGAAGACTTGAGGGCGGA 15143
 QY 181 ValValAsnThrProValSerAsnLeuArgLysGluValArgGluValMet 198
 Db 15144 GTAGTAAATACGCCCCCTTATGTAATTTGCCCAATTAGTAGAGAGAGATTATG 15197
 RESULT 6
 AF190581 15468 bp RNA linear VRL 04-JAN-2000
 LOCUS Beet yellow virus strain BYV-4, complete genome.
 DEFINITION AF190581
 ACCESSION AF190581
 VERSION AF190581.1 GI:6492367
 KEYWORDS
 SOURCE
 ORGANISM
 Beet yellow virus
 Beet yellow virus
 Viruses; ssRNA positive-strand viruses, no DNA stage;
 Closteroviridae; Closterovirus.
 REFERENCE
 1 (bases 1 to 15468)
 Pareyemilov, V.V., Haghiwara, Y., and Dolja, V.V.
 HSB70 homolog functions in cell-to-cell movement of a plant virus
 Proc. Natl. Acad. Sci. U.S.A. 96 (26), 14771-14776 (1999)
 MEDLINE
 JOURNAL PUBMED
 10611288
 REFERENCE
 2 (bases 1 to 15468)
 Pareyemilov, V.V., Haghiwara, Y., Alzhanova, D., and Dolja, V.V.
 Direct Submission
 AUTHORS
 Submitted (27-SEP-1999) Botany and Plant Pathology, Oregon State
 University, 2082 Cordley Hall, Corvallis, OR 97331, USA
 JOURNAL
 Location/Qualifiers
 FEATURES
 source
 1..15468
 /organism="Beet yellow virus"
 /mol_type="genomic RNA"
 /strain="BYV-4"
 /db_xref="taxon:12161"
 /note="California strain; RNA transcript derived from
 this cDNA clone causes formation of red necrotic lesions
 in Claytonia perfoliata and systemic infection in
 Nicotiana benthamiana"
 join(108..7997,7999..9381)
 /note="The polyprotein encompasses domains for papain-like
 leader proteinase, putative methyltransferase, RNA helicase
 and RNA-dependent RNA polymerase (RdRp); ORF1a/b;
 and papain-like leader proteinase is capable of
 autoproteolytic release from the rest of the polyprotein,
 and is a replicational enhancer. RdRp is presumably
 expressed via a +1 ribosomal frameshift"
 /codon_start=1
 /product="p348"
 /protein_id="AA014300.1"
 /db_xref="GI:6492368"

translation="MAFLNVAVPSCAPAFAPAFAPAGSPVPSFCVPSYSDISH
 FRLTSLDFSVRPLFLNARVALRASNDNPSPSLPLGPHATFVLELNGSSAPIS
 RHIDFVARPFSVPTFVLVSLSRTSPRLPALCDFLSCSKGPCVEIASFSTPP
 CLVNCVQAOPTTHAEMESIRPTKTLFAGRLGFCRKLKTYRKEPTLLIHESGLAKLS
 ALGVSKNSRPITVKSASGEKYEYRSRDPFRSRRQOTPRVSGHKPKINKAYE
 PFPPEPRKDKRKRASLPTKDBGFIITGTLRFLSLSRPEPRPLPRFREVEIIVVKK

[illegible]

CDS
 syddavlnsselllgrvipklrsgsrckldv"
 11311..12972
 /note="64 kDa protein; ORF4"
 /codon_start=1
 /product="p64"
 /protein_id="AAFI4303.1"
 /db_xref="GI:6492371"
 /translation="MTRFSPFANYYWGELEFRFRPFGRKWNKINSEASVRRPYORD
 FRPNSGLTSSKNSDSTGESFVEFELLTFPTATYVCVLCCGVAEOLASGMARLYD
 YNVEINLVQKRWGCKGNQIQTVEFPRKNGINAEISLHEGMSLNSGCELIPLFD
 TKRVSLIIFKKQVSTDEAVSVSSVLYDLSHGIN.YETON.SSNGSKALIDEXLK
 YVIVYLEISDIEYRSISDNPLVAGVLYDMCEPYNITLKTSLYKINIESPCELSIATPLL
 SELISWMEQDAPVDVLLFELDTTELILKIPITNTHDSTLYKKKLYLSSTEDBN
 ELIRKVDLSLTRDNPEIKLCAQRWGHCYGVGRIRQTRKVRDAEYKLPALDEGT
 INMGVEEFPDQCKMPSVSVRRFCQSLSEHAFSLFKRGVGPPIITLNNPVKVS
 YLANDYVYHVRAGLITQDELITLISNIEFDVAEMCCEREVALQARPARQGRGEPFGWG
 VKNVSPARASIRVKKSNESLNLVLMKQVGRQRLNPLHRK"
 12970
 /note="ORF5; CPM"
 /codon_start=1
 /product="minor capsid protein"
 /protein_id="AAFI4304.1"
 /db_xref="GI:6492372"
 /translation="MLAPEGKGLDIHFENTRDAMETFFNSYDLAESEVNPINLNRK
 EFDLIGVIRPFRFSELVITDDEPVLKALPALLRAAITSTKVNYSAAVEYITGGKK
 ELVDKAWFPIIKCKMKRKNRPVRFCCATPEAYIVIRASLKPILNITIGRGIP
 SGYEPLGADFIETASVCLINDBERKAIIVQASPAIDRAVSSSDKIVSLIDGLRLS"
 13641..14255
 /note="ORF6; CP"
 /codon_start=1
 /product="major capsid protein"
 /protein_id="AAFI4305.1"
 /db_xref="GI:6492373"
 /translation="WGSABEPIGSAITENNSLADQTLAEGDCKLRDREBECIKLKG
 VPEEDLGLALGCLWYSCATIGTSNKSVAVQPISTIKASFSGKGLFTHGSLRFLDS
 OKLLEGKNGKRCRCRTFOXDYISPAKEYRGLPPIARNRHGLPADDHLLADAFIST
 STELDIDQGRILLARENATHTFSSSPVTSLKQLGRGLATGR"
 14255..14785
 /note="70 kDa protein; ORF7"
 /codon_start=1
 /product="p20"
 /protein_id="AAFI4306.1"
 /db_xref="GI:6492374"
 /translation="MTSSVELAQTKPLFRVILLKGFVFIYVIAIETEESEPEVELPLYV
 LHDLELVNKKGQILSESYIDKSCMTLRKPSVSYTRFASGKPEDFSSSYSGKTFPS
 KILNRKVFTEKSGIOLVFGWGRDQCVSEYILMENVVFAHCGTLPYCLNCELDK
 SGCEELITFEKRNKVFLL"
 14789..15302
 /function="enhancer of RNA accumulation"
 CDS
 CDS

```

3609. 11405
/ncore="HSP70 protein homolog; ORF3, required for
cell-to-cell movement"
/codon_start=1
/product="HSP70h"
/protein_id="AAJ14302.1"
/db_xref="GI:6492370"
/training_set="MVVYVLDLFGFTTSVCAVGGEEYLYFKRDSAYITPTFPHSPDSD
QVAGVYAAEYVLSNDPSVRGSGFYVLDLKRWIGVSEELKRLPHKTLTVAAS
SSTVRLVYSGVTPQNAITLGLATLTKALISTASBAKCOCTEYTSVPANNCIQ
SFSTSCVNLGSGPCVYVNVNPSAAALSTAKGATSLVLDTSKGSFVSYSIAL
NNTFVRAAGSGMNGIDLDKAVELHLYKQALPNVYKIDLTPEKISKVSFSEI
PVASQNTYVVDLVNVSLEAVLAFPERITLKVKEVYKSGSRLEPSYAKLIMW
GGSTLPGILSLTSPVEVECLVLPARAAVGGCAISACLRNDSPLLVDCAANH
LISLKEVCEYVGNCPAGSPITFGSVTPVMAAGSASVAYSALEFGDFYKORLKKR
IGDYVLDGVGVGCTSRVPTLTINBSGCTTTPSLVPTPTKTLVGNAGVAPSSY
QUGERVVLDLHNSDKYLTHALTVPFQKKLTUSDALFLKLSADYRREAKFS

```

ORIGIN	
Alignment Scores:	
Pred. No.:	7.31e-21
Score:	297.00
Percent Similarity:	54.50%
Best Local Similarity:	37.57%
Query Match:	29.97%
DB:	14
	Gaps: 2
US-09-613-486-15 (1-198) x AF190561 (1-15468)	
QY	7 SeranLeuSeranLeuVal11leThrAPAlaseSerLeuAsnG1YValAspLysIys 26
DB	13671 GCAGCTTTGAAAACCGAGCTCCGGAGACCAAACTGTTTGACCGGTGACGCTCGCAC 13730

CDS
 14252..14785
 /function="unknown"
 /note="ORF 7"
 /codon_start=1
 /product="20 kDa protein"
 /protein_id="AAC25121.1"
 /db_xref="GI:3283083"
 /translation="MTSSVELAQTPLFLVLLKGFVYVAIEETSESPVEVLELYV
 LHDLEVMKKKIESYDIFSKSCMTLRKPSSESYTRVSGKPSSEDFSSYSGKFPDS
 KTLNKKMPTEPESGILVFGMGVGRDRCSSSYLMFENVFVGAHGTLPYGLNCELDK
 SGGEELITFSKREVFV"
 14769..15302
 /function="replication enhancer"
 /note="ORF 8"
 /codon_start=1
 /product="21 kDa protein"
 /protein_id="AAC25122.1"
 /db_xref="GI:3283084"
 /translation="MKFFNDCESTRRAISRSLSLRVKEJCTNPSQSEVSECINEFN
 ELARFNHLVTVHEHREMRKHPQSSLETSRLSEMKEIPALIKVAVVPMHEITA
 SETLNALFEYCGITGLITREDALREMRKVRSTVLFHSEELKFEVTNMFETELK
 LNLIRVSISSQILGIIV"

ORIGIN

Alignment Scores:
 Pred. No.: 1,2e-20 Length: 15468
 Score: 295.00 Matches: 71
 Percent Similarity: 53.97% Conservative: 31
 Best Local Similarity: 37.57% Mismatches: 85
 Query Match: 29.77% Indels: 2
 Gaps: 2

US-09-613-486-15 (1-198) x AF056575 (1-15468)

QY 7 SerAsnLeuSerAsnLeuValIleThrAspAlaSerSerLeuAsnGlyValAspIleYs 26
 Db 13671 GCGACTTTGAAAACGAGAGCTCGTAGACCAACGCTGTTGACGCGTAGACGCGAC 13730
 QY 27 LeuLeuSerAlaGluValGluYsMetLeuValGluIleGlyValAProAsnGluGlyLe 46
 Db 13731 AACCTCGGAGAAATTCGAAAGAGCTTTGAAAATTGAAAGGGCTTCGGAAGATAAATC 13790
 QY 47 GluValAlaPheGlyLeuLeuLeuYValAlaLeuAlaAlaArgThrThrSerProYsVal 66
 Db 13791 GGTCTCGCGTTAGACTTTGTTGTACTCTGTCGACGAGTAGACTCTTATAAAGTT 13850
 QY 67 GluArgAlaAspSerAspValIlePheSerAspSerPheGlyGlu---ArgAsnValVal 85
 Db 13851 AGCGTCCAAACGAGCTTACTTATCAACAAAGCTTCGTCGAGTGGAGAAAGATTGTTTC 13910
 QY 86 ValThrGluGluYAspLeuYsValLeuAspGlyCysAlaProLeuThrArgPheThr 105
 Db 13911 CTCTCTACGAGCAACTGAGAGCTCTTCTGACTCTTCAGAAATTTTAGAGGAAAGCCT 13970
 QY 106 AsnYsLeuArgThrPheGlyArgThrPheThrGluAlaIleValAspPheCysIleAla 125
 Db 13971 AACAAATTCGCGTGTCTCTCGCAGACTTTTCAGAGATTATCATCTTCGCGAAGAA 14030
 QY 126 TyrIleHisIleYsLeuProGluLeuAsnAlaAlaGluLeuGlyIleProAlaGluAsp 145
 Db 14031 TACCGAGAAAGACTGCTCCGATTGCTAGAGCCCAACGTCACGCTTACTGCTGAAGAT 14090
 QY 146 SerTyrLeuAlaAlaAspPheLeuGlyThrCysProYsLeuSerGluLeuGlnGlnSer 165
 Db 14091 CACTACTTAGCTGCTGATTCAATCGACATCAACAACTTACTGCTACCAACAAGGT 14150
 QY 166 ArgYsMetPheAlaSerMetTyrAlaLeuYsThrGluGlyValValaAsnThrPro 185
 Db 14151 CGTCTGCTGTTGGCGCGAAACGCACTCACACAGAAATTCG---TCTGATTCACCA 14207
 QY 186 ValSerAsnLeuArgGlnLeuGlyArg 194

Db 14208 GINACTGTTTGAACAGCTGCGTCT 14234

RESULT 9
 BYVCPG 2724 bp ss-RNA linear VRL 02-AUG-1993
 LOCUS Beet yellows virus coat protein gene, complete cds, and two ORFs.
 DEFINITION M59452.1 GI:323237
 VERSION coat protein.
 KEYWORDS Beet yellows virus
 SOURCE Beet yellows virus
 ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage;
 Closteroviridae; Closterovirus.

REFERENCE
 1 (bases 1 to 2724)
 Brumsted, J., Moseley, J. and Hull, R.
 Nucleotide sequence of cDNA encoding the coat protein of beet
 yellows virus
 JOURNAL Unpublished (1990)
 COMMENT coat protein
 FEATRES Original source text: Beet yellows virus cDNA to genomic RNA.
 source location/Qualifiers

CDS

1..2724
 /organism="Beet yellows virus"
 /mol_type="genomic RNA"
 /isolate="Broom's barn"
 /db_xref="taxon:12161"
 <1..1287
 /codon_start=1
 /label=ORF 1
 /protein_id="AA072953.1"
 /db_xref="GI:323238"
 /translation="PDEFVRKLNGLAEPSLVHCHWSLSNCGELINPKDTKRVSLIF
 KGDVVESTDEAVVSSYLIDYLSHCNLVETCNLSNSGKALYDEFLKVIYVLETS
 DLEVRSPDNPVAGVLYDMCFEYNTLSTYLNKRIYSEYFEDSNELIVKDS
 OPADVRLPELDIAELLIKPIINTDSTFLYKNAKRYSEYFEDSNELIVKDS
 LTRDNPBEIKLQMGVGFHCVYGFVRAQTKYRDAEYVLPAAGFTINMGVEEF
 FDELQKMPYSVYRRRCGSLSHASBIFRPGVFPPIRLNVPVYSYLVNDYRH
 VKRAGLQDELTILSNIEFVVAEMCCERFALQRRARQREKRFQCKYKANEVSPIA
 RSSIRVKSNESSLNLTIMKDVGARQRLNPLHKFI"
 1235..1885
 /codon_start=1
 /label=ORF 2
 /protein_id="AA072954.1"
 /db_xref="GI:323239"
 /translation="WLAPEBGRGLIHTENTRDMETFPFNSYDLAESEYNPKLNKX
 ETDLGAVIRERFSELTIVDEDFVKRLARLALAGNITTSYKNTYGAAYETTGKX
 FLVVDAMVFLIKEMCKRPNPVRFCATFEDAVYIVARSLPKLNFTIGRGILP
 SGVEFLGADFLTATSVCLNDHKAIVLAQRAAIDRAVSSSVGKIVSLDLGLS"
 1956..2570
 /codon_start=1
 /label=ORF 3
 /product="coat protein"
 /protein_id="AA072955.1"
 /db_xref="GI:323240"
 /translation="MGSAEPISAIATFENVSLADQTCIAGEDCDLKRNFEECKLKG
 VPEDKIGLGLGYSCTATGTSNKKVVOPTSTIKASFGGKELFTLHGLRFLDS
 OKLEGNKTRRCRTFOKDYISFAKEYGRPLPIRANHGIPADHYLAADPISIT
 STELTDLQGGILLARENATHTTSSSPVSLKQLRGLATGR"

CDS

ORIGIN
 Alignment Scores:
 Pred. No.: 1,99e-21 Length: 2724
 Score: 293.00 Matches: 71
 Percent Similarity: 53.97% Conservative: 31
 Best Local Similarity: 37.57% Mismatches: 85
 Query Match: 29.57% Indels: 2
 Gaps: 2

US-09-613-486-15 (1-198) x BYVCPG (1-2724)

QY 7 SerAsnLeuSerAsnLeuValIleThrAspAlaSerSerLeuAsnGlyValAspIleYs 26
 Db 1986 GCGACTTTGAAAACGAGAGCTCGTAGACCAACGCTGTTGACGCGTAGACGCGAC 2045

CDS
 /note="unnamed protein product; ORF4"
 /codon_start=1
 /protein_id="CAA37552.1"
 /db_xref="GI:58881"
 /db_xref="SPTREMBL:066110"
 /translation="MTTRSTPNNYTWGLFRRFRGGGCKMLMEASVSRPRYSDF
 FRPSDVLIRKTFGSGESFVREPSLLTPPKTEVCKLGVAMETLNNMNLSD
 YNVEFNIVKTVGCKFNIOQVTEFVKINGVAPSPLEHOMSLNCGGLIAPKD
 TKREVSILFKGDLASTDEAIVSSYLDSHCLNLYETCNLSNGSKSLYDFELK
 HVDYLENSDLREKSPDNPLVAGIYDMCFEYNTLSTYLNKLSFDFSLYPL
 SEVPMWNERPAPDVRLPELDAELLVPTIMMDSFTLYKNKLRYLESEYEDSN
 ELIKVDSLLTRDNPDLKLAQWVGHCYGFRTAOTRKXKRAEYELPALBEPV
 LNNVGYRHYKRVGLTQDELTLISNIEFDVAEMCCEREAVALARRAQREKPEQCKG
 TKREISPHARSIRVKKNDLSLITIMKVGASQRLNPLHKKH"
 4185..4835

CDS

/note="unnamed protein product; ORF5"
 /codon_start=1
 /protein_id="CAA37553.1"
 /db_xref="GI:58882"
 /db_xref="GOA:008538"
 /db_xref="SPTREMBL:008538"
 /translation="MLAPEARGLIHENTRTDAMETFFNSYDLAESEVNENKLNK
 ETPDELGVIREKRSKSLVTDQDFVKHLAFLIRANITTSVGVNVAEYETIGGK
 FLVKDAVFPFLICECKKPKPKPVPFCATPEDPAIVARSLPKLFLYRTIGKGI
 SGVFLGADFLTATVCLNDHEKAIYIQAARAIIDAVSSVDGKIVSLFDGRLS"
 4906..5520

CDS

/codon_start=1
 /product="capsid protein"
 /protein_id="CAA37554.1"
 /db_xref="GI:58883"
 /db_xref="GOA:008531"
 /db_xref="SPTREMBL:008531"
 /translation="WCSAPISATIPFENVSLADQTCLEEDCDKLRKPECKLK
 VPEDNLGIALGLCLYSCATIGSNKVNQSTFTFICASFGGKELVTHGELSLFGS
 QKLEGGKPKLRCECFRTFOXDYSLRKEVYKGLPIARARHGLPAEDHYLAADP
 STETLDQGSRLILARENATHTFSSESPTSJKQJRGIGTGR"
 5517..6059

CDS

/note="unnamed protein product; ORF7"
 /codon_start=1
 /protein_id="CAA37555.1"
 /db_xref="GI:58884"
 /db_xref="SPTREMBL:008544"
 /translation="MTSSVELACTKPLFRVILLKGFYVVAETEBESSEALPLVY
 LHDELINRKXKIEASYDPMSCMTLRKSVYTSSEKSEDEFSLPGSGKTS
 KUNRKRTFTRPENGVOIVPGMYGLGQCVSDYIMFENVVGAHCGTLTYCINCELDK
 SGGELEILTFSSKZVLLKRW"
 6034..6567

CDS

/note="unnamed protein product; ORF8"
 /codon_start=1
 /protein_id="CAA37556.1"
 /db_xref="GI:58885"
 /db_xref="SPTREMBL:008545"
 /translation="MKRFIDKNGERSALSSSESLAPRYKELGTNSOOSSEIPEK
 ELASFNLVLEVEHREMEQHPNOSCKLRPSRIGETLRIRAPLRYRYVTPHCKTA
 SDTINAPLEECRTGLAREDALEKRVKVSVALPHSELEKFEVTENMFSTTELIK
 LNTLSIRVISOILGMAT"

ORIGIN

Alignment Scores:
 Pred. No.: 3,72e-20 Length: 6746
 Score: 286.00 Matches: 71
 Percent Similarity: 52.91% Conservative: 29
 Best Local Similarity: 37.57% Mismatches: 87
 Query Match: 28.86% Indels: 2
 DB: 14 Gaps: 2

US-09-613-486-15 (1-198) x CUBV3PH (1-6746)

QY 7 SeranLeuSerAenLeuVal11eThraSpAlaSerSerLeuAnGlyValaAspLysLys 26
 Db 4936 GCGACTTTGAAAACCAATGCTCGCAACCAACCTGTTGACCGAGAAAGCTGGAT 4995
 QY 27 LeuLeuSerAlaGluValaGluLysMetLeuValaGlnLysGlyAlaProAnGluGlyLe 46
 Db 4996 AACCTTGAGAGAACTTCGAAAGCTGTTGAAATTAAGAGGGTCCGGAAGTAACCTC 5055
 QY 47 GluValaValPheGlyLeuLeuLeuTyraLeuAlaIleArgThrThrSerProLysVal 66
 Db 5066 GGAATCGCGTTAGACCTTTGTTTCTATTCCTGCTGCTACGATAGGCACTTCCACAAAGTT 5115
 QY 67 GlnArgAlaAspSerAspVal11ePheSerAnSerPhe--GlyGluAenValaVal 85
 Db 5116 AACGTCACCGACGCTTACCTTCATCAAGCTTGTTGATGATGGAGGAAGAACTGTAC 5175
 QY 86 ValThrGluGlyAspLeuLysLysValLeuAspGlyCysAlaProLeuThrArgPheThr 105
 Db 5176 CTCACTACGCTGAAATTCGATTCCTTCTGGGCTCTCAAAAACCTTGGAGGAAACCT 5235
 QY 106 AsnLysLeuArgThrPheGlyArgThrPheThrGluAlaIleArgThrPheCysIleAla 125
 Db 5236 AACCAATGCGGTGTTCTGCGGTAATTCAGAAAGACTACATATCCTTGGCGAAGAA 5295
 QY 126 TyrLysHisLysLeuProGlnLeuAsnAlaIleAlaGluLeuGlyIleProAlaGluAsp 145
 Db 5296 TACCCAGGGAATTAATCCGATCCAGACCTTAACCGCTACGCTGTAACCGCTGAAGAT 5355
 QY 146 SerTyraLeuAlaAlaAspPheLeuGlyThrCysProLysLeuSerGluLeuGlnIleSer 165
 Db 5356 CACTCTTAGCGCGTGAATTCATTCACGCGTCAAGAACTCACTGACCAACAAAGT 5415
 QY 166 ArgLysMetPheAlaSerMetTyraLeuLysThrGluGlyValaValAsnThrPro 185
 Db 5416 CGTCGCTGTTAGCCCGGAAACGCACTCAACGGAATTCG---TCTGAATCACCG 5472
 QY 186 ValSerAsnLeuArgGlnLeuGlyArg 194
 Db 5473 GTACCACTTGAACCACTAGTGCT 5499

RESULT 12
 BYVUA 15480 bp RNA linear VRL 19-MAY-1995
 LOCUS Beet yellows virus (isolate Ukrainian BYV-U) complete genome.
 DEFINITION X73476
 ACCESSION X73476.1 GI:405624
 VERSION 1
 KEYWORDS coat protein; heat shock 70-related protein; methyltransferase; papain-like protease; RNA dependent RNA polymerase; RNA helicase.
 SOURCE Beet yellows virus
 ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae; Closterovirus.
 REFERENCE 1 Agranovsky, A.A., Koonin, E.V., Boyko, V.P., Mais, E., Frotschl, R., Lunna, N.A. and Atabekov, U.G. Beet yellows closterovirus: complete genome structure and identification of a leader papain-like thiol protease Virology 198 (1), 311-324 (1994)
 JOURNAL MEDLINE 94082464
 PUBMED 8259666
 REFERENCE 2 (bases 1 to 15480)
 Agranovsky, A.A.
 AUTHORS Direct Submission
 TITLE Submitted (18-JUN-1993) A.A. Arganovsky, A.N. Belozersky Institute, Moscow State University, 119839 Moscow, Russia
 COMMENT Related sequences: X53462.
 FEATURES
 source
 1..15480
 /organism="Beet yellows virus"
 /mol_type="genomic RNA"
 /isolate="Ukrainian (BYV-U)"
 /db_xref="taxon:12161"

```

/c1one="DBYV p36, x19, R9, R3, 36A, R2, 36, p43, 111-7,
112, 115, 1213, 1210, 124, 1313, 1311, 142, 143, 1311,
142, 143, 156, 1510, 154, 1518, 1520, 169, 1615, 175"
1..107
join(108..7983,8045..9393)
/gene="ORF1a/b"
join(108..7983,8045..9393)
/gene="ORF1a/b"
/codon_start=1
/product="fusion protein of papin-like protease,
methylintransferase, RNA helicase"
/protein_id="CA51871.1"
/db_xref="GI:809539"
/db_xref="GOA:008534"
/db_xref="SPTREMBL:Q08534"
/translation="MAFLNNAVSCFAPAPAPAHGASPIVPSPPCVRYSDISH
FRLTLDSPVPRPLINARVHLASNDNPLPLPGFAETPLVLELNGSSAPFSIP
RHIDPVNRPSPVPRPVSVSLRPSRLPACLPYCSGBCVCIASGPPPP
CLVSNCAQIPTHAEMESTIRPTKTPAGCFLOFHKRKTTPETLTHESGIALKTS
ALGVTSKPNRPITTVKSASGEKYEAIEISKDERRRRQQTFRVSHKPRKINAVE
PFEPBPKDKRRKRALPTDEGFTTGTLPRLSEPRLEPKREVELPVYKX
HAYPAVSKPRTFRPATGAYVNAQNCSRPNHPILREASYTEFGKMPLORF
MKKEKEYKRSKVSSCSVTKSPLEALAILKNLPQYSNSRLKEVHFIQDDDEI
EYHPRGGKISVILLPKBAGCVTAAPQVHAALTARQDRPVGEILOHFRGGEL
CYLAHALCCALOKRTPREDEPFVGMPTKTFVAKLTTELQPSALKHYRGQVRS
LFHODVASAPSPPIPLRTIGVBEBAEPTISLGHKALBETTERVSJHKDILARS
VERKIDLFQDEIKSLSEKRSVTPPYMGAVOSGLTRAPQPNLSFTVSVDHPAA
ARGVLENETLASMANKSPEDIGGCPFLIKRGSPTYHVRPIYDMKDQRRVRLQ
SKYAVLMTVTEGELDERBAFALDALGCDVVDTRDMQYKFGSSCYGKSNKISI
MLPAFTESGNLSVEMYENMGVNYIKITRASAPSEIRGKTLRYRACTEVYOKL
PRDKTKTFLSGDIYLDPAKFSVRPDVYVNCNVMSKPEWMASTKSKENVY
ISGVNIRHVDHIDKSECPAVMLAVANSRTTTEPLAKNLVYTGDSACEITLFL
FREMRSRAVAEINRSPFKMLKSLISAGLDEPLDNLNQLHLEYSBEVRSIAQNG
EVCNENRNVLTETIAEADRSKIVIEACSGALSVPTOPRGJRGSGRSVGLVNL
VEEVGNLPSVGDVARELVKFTGSDSPFRVAVRDLAEASPFVSVGLAMLR
EAVASPSWADRVSRSVKTFVNRVTRFLNMSAKTLTKKFFPFLSALAKAVV
RKAKVILEAVMEVWESILSDSGEYAVRCCSSVITILLNGRLLPGSSPALIEEV
LIDLATKISIEVLKQISPADSTASSALYRPUTSEILSNRTMGHGIPTKYVLGCF
LPVFRKCVALLCPGDMATYARFLEIGVDLPFLGRSVNISKNYLCVAGLVDSYD
SVVILKSGVAKERYLGFPSKIKNFLNVPKACVVRTSSDLSDEYFCSDESGV
RLRGSSRFTLSRLIDIFNLKSKSLVIEACSAVERIERNMKLYFFPLNSEEERAR
LIRACADFPYLSADDEDEMLRQAEOQYSSDDSVYDGKPTVLRSLNVSERFL
ETFGNCFPVKSVNFKALYSRLRYLPAWDRNLSDPRLKGNENKALKAFKFCV
ITGCEVSOICCLRLTRCWGTBACGIVRLFYITGSTRYLHVVAVALVACPLVNE
LDGSDLTWNGSVPRRLFYALRRLASVNSNLRKTIETPRGNHPRDYAVIET
NEVAPESPVEDIDVDCDFGSDSESVSDVANSRPPGLHGSRSNPLTSLVAY
FKLARIRPLFRINFAVFERRLASRKLKFTIGLARLPDNFSLTSLVYLLQEDVS
VLSAFLIDVELILNCSGVNLPVSWRGSILTALAEIVSGASPLGRNCVCSWDC
SSSNAGCNFMSPVRTKGFVPSSSGSSTASMYERLEAESDIREVLTGRCGSEB
EERKEVTEPGIEHTEDEVPIRSHSGPLSGECSYSEDEENERANLLPHYKYLEE
RSGELTRMRKRTLHGVSEFLNINTSNEQRPPIVDHSPESRALNVSVEFYIOL
ALPELCKLAEEYDOLKVANRROECLCDKDDMPVLRAGGVASGRNRLPKHKXG
HEFRFRSGCLVPYDGTSRVDITPHYQNPVSNALMAGIYSTPTFTTNLSAVLLYE
APRGGKTTLLIKVCEFTSKVNSLITANKSRBELLAKVNRIVLDEGTPQLQTRD
ILITDSYLMNRKGLTCVLYLDCFVYHGAACAEFTKCDASLIFGSRQVRYGRG
SELDTAVLSIDLRVDDSESRVGVGYRCWDVACAMSTPYPTVATVNNVSGQSM
QVRIEISVDVEYSEFVYLTMLQSEKDKLFGKGRSRSEVETVLTVHEQGETY
KRVNLVTRKFOEDPFRSENNHTVALSRHVESLTVSLSKDDALIAQAVKAKQVD
AYRYPTSPGASLTIVSNPSTDRSKCASAPYEYVINSFLESVPGTTSVPGQV
BEMGTQVFBEGADNVIRDSAPVNSKTDHEENLYSVSRPNFTVCERSGPPEQQA
MAWVLRSPFLEVNARSDVIAITKGVRTMKSREBQULASSDLDKPLNDEE
ITTFEKLWKRDARKLDSCLVHPPPAQNIMPRKAVNALFSPCFQEPKRVITVNS
NITFEKEMTNSIAGIKEMVIGSHVYVNGEIDFSPKQDAFIKSPETVYASAGF
DEDLIDVMOGEYSNATITLDQLSFSDVODKRGKSNWTVIGNSITGLISFYATN
RFKALFVSGDLSLFSBPIRNSADWCTELGEGKTLPSVPFCSKFFVMGHYV
FVDPVYVLTGLASKDEVDDEPLFEVFTSRFDLTDLNDRVIELLTHVSKYAYE
SGDYVYALCAIHCRSNFSPFKLVPYKGVVHYGLKLVAKRANCREREKDTAFG
BRFLTLTKLSTLV"
108..7983
/gene="ORF1a/b"
/number=1
7983..8045

```

```

/gene="ORF1a/b"
/note="+1 ribosomal frameshift signal"
7984..8044
/gene="ORF1a/b"
/number=1
8045..9393
/gene="ORF1a/b"
/number=2
9444..9608
/note="unnamed protein product; putative membrane-binding
protein"
/codon_start=1
/protein_id="CA51864.1"
/db_xref="GI:405627"
/db_xref="SPTREMBL:Q08542"
/translation="MDCVLRSYLLARGFLLCFLPLCLVVFIMFVYKQILFRTAQSN
EARNHSTVV"
9608..11404
/codon_start=1
/product="heat shock 70-related protein"
/protein_id="CA51865.1"
/db_xref="GI:405628"
/db_xref="GOA:P37092"
/db_xref="SWISS-PROT:P37092"
/translation="MNVGGLDGTGTFSSVCAVYGEELYLEKORDSAVITYVFLHSOT
QEVAFGYAEVLSNDLSYRGGFYRDLKRMIGCDEENRDYIEKLRPHYKTELLVYAO
SKSVTKLCYGVTPQNAITPLGLATFPRKALITASBAPKQCTGVCSVPANVNCQ
RSFTFSCNVLGCPVNVNMPSSAAALASGRIGAVSPVLYVPGGTPEVSVIISAL
NNTFVYRSGGDMNLGGHIDKAPVLEYNKQALPVNYKIDISLKSLSKVSFLN
PVSQGVGRVVLVNVSLAEVAAFPVETIKIVAEYKXCSMRLEPNYKALLMY
GSSYLPGLLRSLSPVDECLVLPDRAAVAGGALYSKLRNDSFMLVDCAAH
LSISKVGEISVCPVAGSPIPETGRTVNTGNSAVALFEGDPVKRLKRLIF
RGDYLVANVGTGATNRVPLTETINSSVGTISFSLVGPPTGVYKLLIGNAYDPSY
QUGERVVADLKHNSDKRLHAULTYQFPQKSKULDGDKALFLKRLTADVTRERAKRS
SYDAVANSSELGLRILPKILRGSREKLDV"
11310..12971
/note="unnamed protein product; putative heat shock
90-related protein"
/codon_start=1
/protein_id="CA51866.1"
/db_xref="GI:405629"
/db_xref="GOA:Q08541"
/db_xref="SPTREMBL:Q08541"
/translation="VTRFSRPANYWGLFRFPGGQEWNLSEASVSRPIYSSD
PFRSGVLSIKRTGSESTGESFVBEFSLTFPTVEYCKLGVAMETLALNGNRISD
YVNSRPNIVDKYTKGCKENIQSVTFPKYKINVAEPLVHGMCLNSGCELLNPDP
TKRFRSLFKKQDLAESTDEDAIVSSSYDIYSHCLNAYETGNLSNSGKSLYDPEIK
HVIDLBSLDIERSPSDNPVAGLILDMCEBENTLASYTKNLTBSPDCSLTLPIL
SEVFSMMNERAPDRVRLLELDAABELLKVTIMHDSITLYKRLTLYSYPEDDSN
BLIKVYDGLTIRNDPRLKLAQRAWGFQYGVYRTQTRKRVKDAEYKLPALGEEV
INMSVEBEFFELCKMPSIVRRRCFGLSHEAFVFKRGVGFPPITRLNVVKKX
YLVNDVYHVRVGLTQDELITLNIIEEDVLEMCCERVALQAPRAGGGERDFQGMKG
TKNELSPHARSIRYKKNDSLLNIMWDVGARQORRLNPLHRHG"
12919..13569
/codon_start=1
/product="coat protein homolog"
/protein_id="CA51867.1"
/db_xref="GI:405630"
/db_xref="GOA:Q08538"
/db_xref="SPTREMBL:Q08538"
/translation="WLAPEARGGDLIHPEENTRDAMEFPFNSYDLAESEVNPNLNRK
ETDELIGYIRRRKFSELYITDEDPYKHLAPLILAAINTTSYVKNVYGAAYVTIGKK
FLVKQAWVPLIKCKMKFNPNPRTFCATFEDVYITASDGLVLRITRGISRGIP
SGYEPFLGDFLPLTAVSCINLDEKALIVLQASPAALIDRAVSSVDGKIVSLFLGRSLP
13640..14254
/codon_start=1
/product="coat protein"
/protein_id="CA51868.1"
/db_xref="GI:405631"
/db_xref="GOA:Q08531"
/db_xref="SPTREMBL:Q08531"
/translation="WGAAPISAIATFENVSLADQCTLHGDCDKLRKRNFECKLKLG
VPEDNLGIALGLCYSCATIGSNKVNQVPTSTFIKASFGGKELYLTHGELNSFLGS

```

exon
5'UTR
gene
CDS
m1sec_signal

CDS
 ORKLEGRKPKLRCFCRTFOKDYISLRKRYGKLPPIAARNRGLPAEDHYLAADFIIST
 STELTDLQSRLLARENATHTEFSSESPTSIKQIGRIGTGR"
 14251..14793
 /note="unnamed protein product; orf8"
 /codon_start=1
 /protein_id="CA51869.1"
 /db_xref="GI:405632"
 /db_xref="SPTREMBL:008544"
 /translation="MTSSVLAQTKPLFRYLKGFVYVAFEETESSEAEPLVY
 LHDPELNINRKIEASVYDFMSCTMRKRSVYSVRSSEKSSBDPSFGSGKFGS
 KVLNRKVTPEFNGVOLFVGMGLEQRCVSDYLWENFVFGHCTLTLYCLNCELDK
 SGSELETITPSKNEVLKRW"
 14768..15301
 /note="unnamed protein product; orf8"
 /codon_start=1
 /protein_id="CA51870.1"
 /db_xref="GI:405633"
 /db_xref="SPTREMBL:008545"
 /translation="MKPEFLKDGRTSRASRSESLRKYKELGTNSQSEISRCVDFN
 ELASFNHLITVTHREHMEQHPNOSKLRPSIGMKIEIRFLKRYVTPMHEETA
 SDTLNAPLEBYCRITGTAREDALREKRVKSVLVFHSBELFEVTENMFSTELK
 LNLRLRVISQIIIGMAL"
 15299..15480

RESULT 13
 A41914 LOCUS 783 bp DNA linear PAT 05-MAR-1997
 DEFINITION Sequence 1 from Patent WO9428147.
 ACCESSION A41914
 VERSION A41914.1 GI:2297469
 KEYWORDS
 ORGANISM
 Beet yellows virus
 Viruses; ssRNA positive-strand viruses, no DNA stage;
 Closteroviridae; Closterovirus.
 1 (bases 1 to 783)
 Bojsem, K. and Brunstedt, J.
 DNA SEQUENCE COMPRISING AT LEAST TWO COAT PROTEIN GENES
 Patent: WO 9428147-A 1 08-DEC-1994;
 SANDOZ AG (AT)
 Other publication AU 7122694 941220.
 COMMENT
 FEATURES
 source location/Qualifiers
 1..783
 /organism="Beet yellows virus"
 /mol_type="unassigned DNA"
 /db_xref="taxon:12161"

Alignment Scores:

Score: 1,11e-19 Length: 15480
 Percent Similarity: 286.00 Matches: 71
 Best Local Similarity: 52.91% Conservative: 29
 Query Match: 37.57% Mismatches: 87
 DB: 28.86% Indels: 2
 Gaps: 2

US-09-613-486-15 (1-198) x BYVDA (1-15480)

QY 7 Seranleuseranleuvalillethrasplaserleuansnglyvalasplyslys 26
 Db 13670 GCGACTTTTGAACGTAAGCTCGACAGCCAAACCTGTTGCGAGAGAGACTGGCAT 13729
 QY 27 Leuuseralagluvaligluysmerleuvalglnysgialaproansnglylyle 46
 Db 13730 AAACCTTGAGAGAACTTGAAGAGTGTGAAATTAAGGGTCCGGAAGATACCTC 13789
 QY 47 Gluvalvalpheglyleuuleuvalaleualalaargthrthrseryprolyval 66
 Db 13790 GGAATCGGTAGACTTGTGTTGATTCCTGCTAGTAGGCACTTCCAAAGAT 13849
 QY 67 Gluarglaaspsersaspvalillephesersanserphe--glygluargasnval 85
 Db 13850 AACGTCACCGAGCTCTACTTCAATCAAGCTTGTGGTGGGAAGAACTGTAC 13909
 QY 86 Valthrngluaspyleuyslyvalleuaspglycyalaaproleuthrargpther 105
 Db 13910 CTCACCTACCGGTGAATTCCTTCTGGGGTCTCAAAAACCTTTGGAGGAAACCT 13969
 QY 106 Asnysleuargthrphesglyargthrphethrguialayvalaspphecyserlealea 125
 Db 13970 AACCAATGGCGGTGTTTGTCCGACTTTTCAGAAAGCATACATATCTTGGCGAAGAA 14029
 QY 126 Tyrlyshlslyleuoproglinleuasnalaalagluenglyleproalagluasp 145
 Db 14030 TACCGAGGAATTAACCTCCGATCCGAGACTAAGCTACGCGTCTACCGGTGAAGAT 14089
 QY 146 Serlytleualalaasppheleuglythrscysprolyslseuserguluenglnser 165
 Db 14090 CACTACTTAGCGCGTGAATTCATCGACGTGACGAGAACTCACTCACTCAACCAAGT 14149
 QY 166 Arglysmecpnealasermettyralaleuysthrngluiglyvalaalsnthrpro 185
 Db 14150 CGTGTGCTGTAGCGCGGAAAGCCCACTCACAGGAAATTCCTCG--TCTGAATCACCG 14206
 QY 186 Valseranleuarglnleuglyarg 194
 Db 14207 GTAACCACTTGAACCACTAGGTCT 14233

Alignment Scores:

Score: 2,78e-21 Length: 783
 Percent Similarity: 285.00 Matches: 70
 Best Local Similarity: 53.44% Conservative: 31
 Query Match: 37.04% Mismatches: 86
 DB: 28.76% Indels: 2
 Gaps: 2

US-09-613-486-15 (1-198) x A41914 (1-783)

QY 7 Seranleuseranleuvalillethrasplaserleuansnglyvalasplyslys 26
 Db 119 GCGACTTTTGAACGTAAGCTCGACAGCCAAACCTGTTGCGAGAGAGACTGGCAC 178
 QY 27 Leuuseralagluvaligluysmerleuvalglnysgialaproansnglylyle 46
 Db 179 AAACCTTGAGAGAACTTGAAGAGTGTGAAATTAAGGGTCCGGAAGATACCTC 238
 QY 47 Gluvalvalpheglyleuuleuvalaleualalaargthrthrseryprolyval 66
 Db 239 GGTCTCGGCTTGAACCTTGTGTTGATTCCTGTGCGAGATAGTACTTCAATAAAGTT 298
 QY 67 Gluarglaaspsersaspvalillephesersanserphe--glygluargasnval 85
 Db 299 AGTGTCCAACCGAGCTCTACTTCAATCAAGCTTGTGGTGGGAAGAAATGTTTC 358
 QY 86 Valthrngluaspyleuyslyvalleuaspglycyalaaproleuthrargpther 105
 Db 359 CTCACCTACCGGTGAATTCCTTCTGGGCTCTCAAAAACCTTTGGAAGAAAGCT 418
 QY 106 Asnysleuargthrphesglyargthrphethrguialayvalaspphecyserlealea 125
 Db 419 AACCAATGGCGGTGTTTGTCCGACTTTTCAGAAAGCATACATATCTTGGCGAAGAA 478
 QY 126 Tyrlyshlslyleuoproglinleuasnalaalagluenglyleproalagluasp 145
 Db 479 TACCGAGGAAGACTCGTCCGATTCCTAGAGCAACCGTCAAGCTCTACGCTGAAGAT 538
 QY 146 Serlytleualalaasppheleuglythrscysprolyslseuserguluenglnser 165
 Db 539 CACTACTTAGCGCGTGAATTCATCGACGTGACGAGAACTCACTCACTCAACCAAGT 598
 QY 166 Arglysmecpnealasermettyralaleuysthrngluiglyvalaalsnthrpro 185
 Db 599 CGTGTGCTGTAGCGCGGAAAGCCCACTCACAGAGTCTCG--TCTGAATCACCA 655
 QY 186 Valseranleuarglnleuglyarg 194
 Db 14207 GTAACCACTTGAACCACTAGGTCT 14233

CDS
 7832..8515
 /function="unknown"
 /note="ORF6; similar to analogous BYSV coat proteins of other closteroviruses"
 /codon_start=1
 /product="p25"
 /protein_id="AAC55664.1"
 /db_xref="GI:1388135"
 /translation="MPPQGAELVEHNANKSSLEVSSESTREKVGKFNEDHKTPEKQVNNLNDELREVLGKTELKTLKALDEDIYHVAFFLPAASVSTSPVYKSGSVYSDIRKXTVDAMIPOVKILASKNPGLRAFCASLGMYSVALEPDPFGTBSYKRGAPSGSESYLGADPFTSTCPMSHDAVVALASARNLDSASAQIDKQKVSLEYGRKVVYT"
 misc_feature
 8551
 /note="transcription start site for coat protein subgenomic RNA"
 8617..9264
 /note="Cp; ORF7; similar to coat proteins of other closteroviruses"
 /codon_start=1
 /evidence="experimental"
 /product="coat protein"
 /protein_id="AAC55665.1"
 /db_xref="GI:1388136"
 /translation="MAGNDEGSDDSASQTMKIMFAPFENFARASATCINGENKKTLEERSVVKQDVTESGIPITIGWTLVALATLSTSSKIDIEDKTPVASKIDAVNVTTEEDIKNPVNSLTLLKNYKTKLRPARTFPEERYLRFVROYKHLIPINARAKHGI PADYSYLAADPVQTSNLLKEHQAVLIEGNNAATASSGTTRESAVNLKVLGSSK"
 9261..9743
 /function="unknown"
 /note="ORF8"
 /codon_start=1
 /product="p18"
 /protein_id="AAC55666.1"
 /db_xref="GI:1388137"
 /translation="MNVIESLQFVTPVHEKALVLTDDAFIFPHGVQSHSPYVLWYHGEGAMADGVSFTYKELQDVLNLQYRRMLNVRKTSLSLELIVRATKNAV SIFRCNAPTLIFVTSKGPESMDALVNGIIGPVSVDPTLLSIVESYGNITHPALE"
 CDS
 9746..10306
 /function="unknown"
 /note="ORF9; similar to the 20-kDa protein of CTV and to the 21-kDa protein of BYV, two other closteroviruses"
 /codon_start=1
 /product="p22"
 /protein_id="AAC55667.1"
 /db_xref="GI:1388138"
 /translation="MKLLSDSYVDSINLANKCLNANKSVPLSCIASYNELLIKVMASQSDAIKKNENKLENPFDSDVTDTLEMLRDRQLRAINKEPFCSSNODVKNITRNNTNHSNREPRVFEIQDSFKRGVITSLONLSYKREVTITNFKCA LLDVGMGVSKTFEENYIGENLSEFL"
 10304..10545
 /note="similar to 3' UTR's of CTV and BYV"
 3'UTR
 ORIGIN
 Alignment Scores:
 Pred. No.: 8,38e-18 Length: 10545
 Score: 266.50 Matches: 66
 Percent Similarity: 48.09% Conservative: 22
 Best Local Similarity: 36.07% Mismatches: 94
 Query Match: 26.89% Indels: 1
 DB: 14 Gaps: 1
 US-09-613-486-15 (1-198) x BYU51931 (1-10545)
 Qy 11 AenleuValilleThAspAlaseSerleuAenglyValAspLyLyLeuLeuSerAla 30
 Db 8704 AATTTCCGCGCGCGCGCGAGCTGTTGATGCTGAACAACAGAGAGAGTTGTCAG 8763
 Qy 31 GIUValGluLysMetLeuValGlnLysGlyAlaPheAenGluGlyLleGlyValValPhe 50
 Db 8764 GAATTTCTGTGAGAGTGAAGCTCAGATGTCACCTGAATCCGCACTTCCCAACACTTG 8823

Qy 51 GlyLeuLeuLeuTyAlaLeuAlaAlaArgThrThSerProLySValGlnArgAlaAsp 70
 Db 8824 GGAATGACGCTGTCGCGCTTACCTTACGACATTCCTTAATTCATATGATGAGAT 8883
 Qy 71 SerAspValillePheSerAenSerPheGlyGluArgAenValValThrGluGlyAsp 90
 Db 8884 AAGACACGTTAGTGAACGCTTAAGATAGTCGCTCACTGACGATTAACCTTACGAGAC 8943
 Qy 91 LeuLySValLysLeuAspGlyCySAlaProLeuThrArgPheThrAenLySLeuArgThr 110
 Db 8944 ATTAAGAACTTCGTAACCTCCTTGACCTGCTTAAGACCTAATAGATTAACCTCCGAGTC 9003
 Qy 111 PheGlyArgThrPheThrGluAlaTyValAspPheCySAlaTyLySValSLeu 130
 Db 9004 TTGCTGCGACCTTTTGAGAGAGACTACCTACGCTTGTGAGGCAATCAACACATCCTC 9063
 Qy 131 ProGlnLeuAenAlaAlaAlaGluLeuGlyTlleProAlaGluAspSerTyLeuAlaAla 150
 Db 9064 CCGAATATCGCAAGACGCAACAAACAGGTATTCCTCGCTGATCTGATTTTGGACCT 9123
 Qy 151 AspPheLeuGlyThrCySProLySLeuSerGluLeuGlnGlnSerArgLySMetPheAla 170
 Db 9124 GACTTCGTCGAACACTGACCACTTATTGAAGACGACGACGACGACGACGACGACGACGAC 9180
 Qy 171 SerMetTyAlaLeuLyThrGluGlyGlyValValAenThrProValSerAspLeuArg 190
 Db 9181 GGTAGAAAGCTGCTACGACATCTTCGCGACCACTAGGAGCTCGCGGTACACTTAAA 9240
 Qy 191 GlnLeuGly 193
 Db 9241 TATCTTGGG 9249
 RESULT 15
 AF501867 672 bp RNA linear VR1-02-APR-2003
 LOCUS Citrus tristezza virus coat protein gene, complete cds.
 DEFINITION
 ACCESSION AF501867
 VERSION AF501867.1 GI:20454248
 KEYWORDS
 SOURCE
 ORGANISM
 Citrus tristezza virus
 Citrus tristezza virus
 Viruses; ssRNA positive-strand viruses, no DNA stage;
 Closteroviridae; Closterovirus.
 REFERENCE
 AUTHORS Roy, A., Ramachandran, P. and Brilansky, R. H.
 TITLE Grouping and comparison of Indian citrus tristezza virus isolates based on coat protein gene sequences and restriction analysis patterns
 JOURNAL Arch. Virol. 148 (4), 707-722 (2003)
 MEDLINE 12548851
 PUBMED 12664295
 REFERENCE
 AUTHORS Roy, A., Ramachandran, P. and Brilansky, R. H.
 TITLE Direct Submission
 JOURNAL Submitted (12-APR-2002) Plant Pathology, University of Florida, 700 Experiment Station Road, CREC, Lake Alfred, FL 33850, USA
 FEATURES
 source
 1..672
 /organism="Citrus tristezza virus"
 /mol_type="genomic RNA"
 /db_xref="taxon:12162"
 /country="India: Bangalore"
 1..672
 /codon_start=1
 /product="coat protein"
 /protein_id="AAW22216.1"
 /db_xref="GI:20454249"
 /translation="NDPERKTLKNNKKEKREDDVVAESSRGNLHIDPTLIANDVROLGTOQNALNRDLFTLTKSKYRPLRPEDDQPHIAMLVYLAAYKSSLSQDDPTGVTYTRGVAVDLSDKMTVDVVRNSKIGRTNALKRWGATNNAIVIAFORORNSISYGRPLDAGIIPAGHYILCAEDFLTGAGLTDECAVYIQAKELLKRGADDEVVTVNVAQSGKEFTR"
 CDS

Fri Apr 9 15:30:05 2004

us-09-613-486-15.p2n.rge

Page 19

ORIGIN

Alignment Scores:	
Pred. No.:	9.29e-05
Score:	130.50
Percent Similarity:	40.00%
Best Local Similarity:	27.73%
Query Match:	13.17%
DB:	14
Length:	67
Matches:	27
Conservative:	61
Mismatches:	75
Indels:	57
Gaps:	11

US-09-613-486-15 (1-198) X AF501867 (1-672)

OY	MetGileuMeSerSerAspSerAsnMetLeuSerAsnLeuValIleThrAspAlaSerSerLeu	20
Db	ATGAACCTTACACATCGATCCGACTCTG-----ATAGCAGTGAACGATGCGCCGACTGG	144
OY	21 AsnGlyValAspLysLysLeuLeuSerAlaGluValGlnLysMetLeuValGlnLysGly	40
Db	145 GGAACCCACACAGAACCGCCCTTTGAACAGAGATTGG-----TTTCTTACTCGAAAGGG	198
OY	41 -----AlaProAsnGlnGlyIleGlnValValPheGlyLeuLeuLeuTyrAla	56
Db	199 AAGATGCTTAACTTGCTGCTGACGAGTAAAGACTTCACATAGCATGATGATTATATCGT	258
OY	57 LeuAlaAlaLarghrThrSer	63
Db	259 TTAGCGGTTTAAAGATTCACTGCTTGGCAMAAGTAGATGACACCAGCGCATTAACATACT	318
OY	64 -----ProLysValGlnArgAlaAsp-----SerAspValIlePhe--Ser	76
Db	319 CGGAGGCGTGTGAAGCGATTGCTGTCTGTGCACAAGCTTTGAGACTGACGTCGTTTAACCTCC	378
OY	77 AsnSerPheGlyGlnArgAsnValValValThrGlnGlyAspLeuLysValLeuAsp	96
Db	379 AAGGGTATGTGTAAACCT-----	396
OY	97 GlyCysAlaProLeuThrArgPheThrAsnLysLeuArgThrPheGlyArgThrPheThr	116
Db	397 -----ACTAATGCGCTTCGATGCTGGGGTGAAGACCTTAACGAT	432
OY	117 GlnAlaTyrValAspPheCysIleAlaTyrLysHisLysLeuProGlnLeuAsnAlaAla	136
Db	433 GCCCTTATTATTAAGTTCTGT---AAGACAGATCCGATTGAGTTATGCGAGACGTCGG	489
OY	137 AlaGluLeuGlyIleProAlaGluAspSerTyrLeuAlaAlaAspPheLeuGlyThrCys	156
Db	490 CTAACTACAGAGATTCGCGCTGGATATCATTAACCTATGTGCAGATTTCCTG---ACCGGA	546
OY	157 ProLysLeuSerGlnLeuGlnGlnSerArgLysMetPheAlaSerMetTyrAlaLeuLys	176
Db	547 GCTGGCTTGACTGATTTAAGATGCGCTGTGCATCAACAGTAAAGAACATTTGTGAAG	606
OY	177 ThrGlnGlyGly-----ValIaAsnThrProValSerAsnLeuArgIleGlyArg	194
Db	607 AAGCGAGGGCGTGAAGTCGTA-----GTTACTAATGTCAAGGACGCTTGGGAAA	655

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: April 7, 2004, 01:14:14 ; Search time 306 Seconds
(without alignments)
2748.834 Million cell updates/sec

Title: US-09-613-486-15
Perfect score: 991
Sequence: 1 MELMSDSNLSNLVITDASSL.....GGVNTNPVSNALQGRREV 198

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xjh
-Q/cgnt_1/USPRO.spool/US0961386/runtat.05042004.15120.28252/app_query.fasta_1.391
-DB=N:Geneseq.29Jana04 -OPT=fastcap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOFCLE=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0961386 @CGN.1.1.352 @runtat.05042004.15120.28252 -NCPU=6 -ICPU=3
-NO.MMAP -IARGCOTRY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MAXN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq.29Jana04:*

1: Geneseqn1980s:*\n2: Geneseqn1990s:*\n3: Geneseqn2000s:*\n4: Geneseqn2001s:*\n5: Geneseqn2002s:*\n6: Geneseqn2003as:*\n7: Geneseqn2003bs:*\n8: Geneseqn2003cs:*\n9: Geneseqn2003ds:*\n10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	991	100.0	597	2 AAV08870 Grapevine
2	991	100.0	15500	2 AAV08874 Grapevine
3	286	28.9	688	2 AAQ87853 Sugar beet
4	285	28.8	783	2 AAT42975 BYV coat
5	89.5	9.0	1500	6 ABQ76539 C. albica
6	86	8.7	1197	3 AAC45864 Arabidops
7	86	8.7	1199	3 AAC39012 Arabidops
8	86	8.7	349980	5 AAF86431 Pyrococcus

9	85	8.6	4467	2 AAT68648
10	85	8.6	5311	9 ADD29639
11	84.5	8.5	4766	4 AAA91259
12	83.5	8.4	1266	7 ACA29885
13	83.5	8.4	110000	6 ABA92787-4
14	82.5	8.3	1773	3 ACA47601
15	81.5	8.2	1209	3 AAA72313
16	81.5	8.2	1881	3 AAA72318
17	81.5	8.2	2040	7 AAA50205
18	81.5	8.2	2442	7 ACA23154
19	81.5	8.2	2814	3 AAA72317
20	81.5	8.2	110000	6 ABA90521-09
21	81.5	8.2	349980	6 ABQ81849
22	80.5	8.1	10709	4 AAF28522
23	80	8.1	2133	6 ABA70626
24	80	8.1	2136	7 ACA50630
25	80	8.1	2187	6 ABA90521-09
26	80	8.1	110000	6 ACA33088
27	79	8.0	1353	7 AAF52987
28	79	8.0	1356	4 AAT72221
29	79	8.0	1434	2 ACA43662
30	79	8.0	1434	7 ABL02037
31	79	8.0	2572	4 ABL02036
32	79	8.0	4827	4 ABL02036
33	79	8.0	11597	6 ABA90521-10
34	79	8.0	11597	2 ABA90521-10
35	79	8.0	17919	3 ABA90521-10
36	79	8.0	61313	4 ABA90521-10
37	79	8.0	61313	7 ABA90521-10
38	79	8.0	110000	2 ABA90521-10
39	79	8.0	110000	2 ABA90521-10
40	78.5	7.9	2282	4 ABA90521-10
41	78.5	7.9	110000	6 ABA90521-10
42	77.5	7.8	665	6 ABA90521-10
43	77.5	7.8	14494	4 ABA90521-10
44	77	7.8	912	4 ABA90521-10
45	77	7.8	929	4 ABA90521-10

ALIGNMENTS

RESULT 1

AAV08870 standard; CDNA; 597 BP.

ID	AAV08870	standard; CDNA; 597 BP.
XX	AAV08870;	
AC	AAV08870;	
XX	27-AUG-2003 (revised)	
DT	27-AUG-2003 (revised)	
DT	29-MAR-1999 (first entry)	
XX		
DE	Grapevine leafroll virus type 2 coat protein ORF6 product.	
XX		
KW	GRLAV-2; closterovirus; grape; tobacco; transgenic plant; disease resistance; virus resistance; beet yellows virus; coat protein; ss.	
KM		
XX		
OS	Grapevine leafroll virus.	
XX		
PN	WO9653055-A1	
XX		
PD	26-NOV-1998.	
XX		
PF	20-MAY-1998; 98WC-US010313.	
XX		
PR	20-MAY-1997; 97US-0047194P.	
XX		
PA	(CORR) CORNELL RES FOUND INC.	
XX		
PI	Zhu H, Ling K, Goncalves D;	
XX		
XX	WPI; 1999-045307/04.	
DR	P-RSDB; AAW73482.	
XX		

PI Zhu H, Ling K, Gonsalves D;
 XX WPI: 1999-045307/04.
 DR P-PSDB; AAW73476, AAW73477, AAW73478, AAW73479, AAW73480, AAW73481,
 DR AAW73482, AAW73483, AAW73484.
 XX Grapevine leafroll virus (type 2) proteins and polypeptides - and
 PT encoding DNA, useful e.g. to impart grapevine leafroll resistance to
 PT grape and tobacco plants and detect grapevine leafroll virus.
 XX
 PS Disclosure; Page 12-20; 151pp; English.
 XX
 CC A total of 15,500 bp of the RNA genome of grapevine leafroll virus type 2
 CC (GLRAV-2) has been sequenced and cloned from GLRAV-2 isolated from
 CC infected Vitis vinifera cv. Pinot Noir. About 85% of the total RNA genome
 CC was revealed from 2 different clones. The sequence in the coat protein
 CC gene region was determined and confirmed from several overlapping clones.
 CC The genomic organisation of GLRAV-2 includes 9 open reading frames (see
 CC also AAV08864-72) encoding a polyprotein (see AAW73476) having papain-
 CC like protease, methyltransferase and helicase motifs, an RNA-dependent
 CC RNA polymerase, heat shock proteins, coat proteins and other proteins of
 CC unknown function (see AAW73476-84). These proteins can be used to produce
 CC antibodies, useful to detect GLRAV in samples e.g. by ELISA (claimed).
 CC The nucleic acid molecules can be used to produce probes and primers for
 CC such detection, and to transform host cells (especially Agrobacterium
 CC vitis, Agrobacterium tumefaciens, grapes, citrus, beet or tobacco cells)
 CC and produce transgenic plants (claimed). In particular, they can be used
 CC to impart GLRAV-2 resistance to Vitis scion or rootstock cultivars or
 CC to impart GLRAV-2 resistance to Vitis scion or rootstock cultivars or
 CC Nicotiana (claimed). Because extensive similarity exists between hsp70-
 CC related sequence regions of GLRAV-2 and other closteroviruses, the DNA
 CC may also be used to impart beet yellows virus resistance to beet
 CC cultivars or tristea virus resistance to citrus scion cultivar/rootstock
 CC cultivars (claimed). (Updated on 27-AUG-2003 to correct OS field.)
 CC
 XX
 XX Sequence 15500 BP; 3984 A; 3178 C; 3970 G; 4368 T; 0 U; 0 Other;
 SQ
 Alignment Scores:
 Pred. No.: 1.03e-113 Length: 15500
 Score: 991.00 Matches: 198
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0
 US-09-613-486-15 (1-198) x AAV08874 (1-15500)
 QY 1 MetGluLeuMetSerAspSerAsnLeuSerAsnLeuValIleThrAspAlaSerSerLeu 20
 DB 13584 ATGGAGTGTGATGTCGACGACGACCTTACGACCTGGTATACCGACGCGCTTACGCTA 13643
 QY 21 AsnGlyValAspIleValLeuLeuSerAlaGluValGluValGluValGluValGluVal 40
 DB 13644 AATGCTGTCGACGACGACCTTTATCTGCGAAGTTGAAAAAATGTTGGCGCAAGGG 13703
 QY 41 AlaProAsnGluGlyIleGluValValPheGlyLeuLeuLeuValAlaAlaAlaArg 60
 DB 13704 GCTCTTAAACGAGGATTAAGAGTGTGTCGCTTATCTCTTAAACGACCTCGGCGCAAG 13763
 QY 61 ThrThrSerProIleValGlnAlaAlaAspSerAspValIlePheSerAsnSerPheGly 80
 DB 13764 ACCACGCTCTCTAGAGTTACGCGCGGACATTCAACGCTTATTTTAAATAGTTTCGGA 13823
 QY 81 GluArgAsnValValValThrGluGlyAspLeuValValLeuAspGlyCysAlaPro 100
 DB 13824 GAGAGGAAATGGAGTAACGAGGGTGACCTTAAAGAGTACTCGACGGGTGCGCCT 13883
 QY 101 LeuThrArgPheThrAsnIleValLeuArgThrPheGlyValGlnThrPheThrGluAlaTyrVal 120
 DB 13884 CTCACCTAGGTCACCTAAATPAACTTAGAACCTTGCTGCTACTTCACTGAGGCTTACGTT 13943
 QY 121 AspPheCysIleAlaTyrIleValHisIleValLeuProGluLeuAsnAlaAlaIleGluLeuGly 140
 DB 13944 GACTTTGTATCGGTATPACCAAAATTAACCCCACTCAACGCGCGCGGCAATTTGGGG 14003

QY 141 IleProAlaGluAspSerTyrLeuAlaAlaAspPheLeuGlyThrCysProIleLeuSer 160
 DB 14004 ATTCACGCTAAAGATTCCTTAACTGTCGACGATTTTCTGGGTCTTCCCGAAGCTCTCT 14063
 QY 161 GluLeuGlnIleSerArgIleMetPheAlaSerMetTyrAlaLeuIleThrGluGlyGly 180
 DB 14064 GAATTACAGCAAGATGAGAGATGTTCCGACGATATGATACCTCTAAAACTGAAAGGTGGA 14123
 QY 181 ValValAsnThrProValSerAsnLeuArgGluLeuGlyValArgGluValMet 198
 DB 14124 GTGGTAAATACACCGAGCGAATCTGCGTACGCTAGGTGAGGAGAGTTATG 14177
 RESULT 3
 ID AAQ87853 standard; cDNA, 688 BP.
 AC AAQ87853:
 XX 16-OCT-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 02-NOV-1995 (first entry)
 XX
 DE Sugar beet yellows virus capsid protein coding gene.
 XX
 KM Sugar beet yellows virus capsid protein; transgenic plant; ss.
 XX
 OS Beet yellows virus.
 XX
 FH Key location/Qualifiers
 FT CDS 74..688
 FT /*tag= a
 FT /product= "sugar beet yellows virus capsid protein"
 FT
 XX
 PN RU2017820-C1.
 XX
 PD 15-AUG-1994.
 XX
 PF 27-JUN-1991; 91SU-04950054.
 XX
 PR 27-JUN-1991; 91SU-04950054.
 XX
 PA (IMMU=) IMMUNOBIOTECHN INST.
 PA (BIOT=) BIOYRCHN INST CO LTD.
 XX
 XX Arganovskii MA, Bolko VP, Karasev AV;
 FI
 XX WPI: 1995-113715/15.
 DR P-PSDB; AAR72682.
 DR
 XX
 PT Sugar beet yellows virus cDNA fragment encoding capsid protein - useful
 PT for production of virus-resistant transgenic plants.
 PS
 Claim 1, Col 7-10; 5pp; Russian.
 XX
 CC The nucleotide sequence of the novel gene encoding the sugar beet yellows
 CC virus (SBYV) capsid protein. The gene encodes a protein of 204 amino
 CC acids with mol. wt. 22.2 kD. The gene was obtained from reverse
 CC transcribed RNA isolated from purified SBYV. The fragment is useful for
 CC the production of virus resistant transgenic plants by genetic
 CC engineering methods. (Updated on 25-MAR-2003 to correct PN field.)
 CC (Updated on 25-MAR-2003 to correct PF field.) (Updated on 16-OCT-2003 to
 CC standardise OS field)
 CC
 XX
 SQ Sequence 688 BP; 190 A; 155 C; 157 G; 186 T; 0 U; 0 Other;
 XX
 Alignment Scores:
 Pred. No.: 2.42e-26 Length: 688
 Score: 286.00 Matches: 71
 Percent Similarity: 52.91% Conservative: 29
 Best Local Similarity: 37.57% Mismatches: 87
 Query Match: 28.86% Indels: 2
 DB: 2 Gaps: 2

US-09-613-486-15 (1-198) x AAO87853 (1-688)

QY 7 SerAsnLeuSerAsnLeuValIleThrAspAlaSerSerLeuAsnGlyValAspLys 26
 Db 104 GCGACTTTGAAAGCGTAAGTCTCGACAGCAACCGTTTGACGAGAGAACTGCGCAT 163
 QY 27 LeuLeuSerAlaGluValGluLysMetLeuValGlnLysGlyValAProLengluGlyLe 46
 Db 164 AACCTTAGGAAGAACTCGAAGAGTGTGAAATTAAGAGGGGTCCGAGAACTACCTC 223
 QY 47 GluValValPheGlyLeuLeuLeuTyralaLeuAlaIleThrHisSerProLysVal 66
 Db 224 GGAATCCCGCTAGACCTTTGTTGTAATCTCGTGCCTACGATAGGACCTTCCAAACAAAGTT 283
 QY 67 GlnArgAlaAspSerAspValIlePheSerAsnSerPhe---GlyGluArgAsnValVal 85
 Db 284 AACGTCACACCGACGCTACCTTCATCAAGCTTCGTTGGTGGTGGAGAGAACTGTAC 343
 QY 86 ValThrGluGlyAspLeuLysLysValLeuAspGlyCysAlaProLeuThrArgPheThr 105
 Db 344 CTCACCTACCGGTGAATGTAATCTCTTCTGGGCTCAAAAACCTTTGGAGGAGAAACCT 403
 QY 106 AsnLysLeuArgThrPheGlyArgThrPheThrGluAlaIleValAspPheCysIleAla 125
 Db 404 AACAAATGGCGGTCTTCTGCGCTACTTTTCAGAAAGACATATCTCTGCGCAAGAA 463
 QY 126 TyrLysHisLysLeuProGlnLeuAsnAlaIleGluLeuGlyLeuProAlaGluAsp 145
 Db 464 TACCGAGGAAATTAATCTCCGATTCAGAGCTACCGTACCGTCTACCCGCTGAAGAT 523
 QY 146 SerTyrLeuAlaIleAspPheLeuGlyThrCysProLysLeuSerGluLeuGlnGlnSer 165
 Db 524 CACTACTTACCGCTGACTCATATCGACGCGACGACGACGACGACGACGACGACGACGAC 583
 QY 166 ArgLysMetPheAlaSerMetTyrAlaLeuLysThrGluGlyValValAsnThrPro 185
 Db 584 CGTCTGCTGTAAGCGCGCAAGACGCTCAACGAGAAATCTCTG---TCTGAATCTACCG 640
 QY 186 ValSerAsnLeuArgGlnLeuGlyArg 194
 Db 641 GTAACCACTTGAACCAACTAGTCTGT 667
 RESULT 4
 ID AAT42975 standard; cDNA, 783 BP.
 AC AAT42975;
 XX 16-OCT-2003 (revised)
 DT 14-JAN-1997 (first entry)
 XX BYV coat protein cDNA.
 DE
 XX Tobacco mosaic virus; coat protein; beet yellow virus; BYV;
 KM beet western yellow virus; BMV; beet necrotic yellow vein virus; BNBYV;
 KM tobacco etch virus; sugarbeet; fruit; mango; apple; pear; banana;
 KM field crop; sunflower; wheat; barley; maize; vegetable; potato; carrot;
 KM cabbage; onion; ss.
 XX Beet yellows virus.
 OS
 XX Key Location/Qualifiers
 FH sig_peptide 13..78
 FT /*tag= a
 FT /note= "Omega sequence from TMV"
 FT CDS 89..703
 FT /*tag= b
 FT /product= "BYV_coat_protein"
 XX MO9428147-A1.
 XX 08-DEC-1994.

XX 01-JUN-1994; 94WO-EP001786.
 PF 02-JUN-1993; 93GB-00011332.
 XX (SANO) SANDOZ LTD.
 PA (SANO) SANDOZ PATENT GMBH.
 PA (SANO) SANDOZ-ERFINDUNGEN VERM GES MBH.
 PI Bojken K, Brunstedt J;
 DR WPI, 1995-022818/03.
 XX Recombinant DNA sequence comprising at least two coat protein genes -
 PT used to combat viruses in e.g. sugar beet.
 XX Disclosure; Page 11-12; 23pp; English.
 PS The sequences given in AAT42975-77 represent cDNA sequences encoding coat
 CC proteins from beet yellow virus (BYV), beet western yellow virus (BMV)
 CC and beet necrotic yellow vein virus (BNBYV) respectively. These sequences
 CC are used in the recombinant DNA sequence of the invention. This comprises
 CC nucleotide sequences encoding the coat proteins of at least one
 CC tobacco virus and at least one closterovirus. This recombinant DNA may be
 CC used to transform sugar beet for combating viral infection. Other plants
 CC may also be transformed e.g. fruit such as mangoes, apples, pears, and
 CC bananas, and field crops such as sunflower, wheat, barley, maize, and
 CC vegetables such as potatoes, carrots, cabbage and onion. (Updated on 16-
 CC OCT-2003 to standardise OS field)
 XX
 SQ Sequence 783 BP; 223 A; 178 C; 171 G; 211 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 3,896-26 Length: 783
 Score: 285.00 Matches: 70
 Percent Similarity: 53.44% Conservative: 31
 Best Local Similarity: 37.04% Mismatches: 86
 Query Match: 28.76% Indels: 2
 DB: Gaps: 2
 US-09-613-486-15 (1-198) x AAT42975 (1-783)
 QY 7 SerAsnLeuSerAsnLeuValIleThrAspAlaSerSerLeuAsnGlyValAspLys 26
 Db 119 GCGACTTTGAAAGCGTAAGTCTCGACAGCAACCGTTTGACGAGGAGAACTGCGAC 178
 QY 27 LeuLeuSerAlaGluValGluLysMetLeuValGlnLysGlyValAProLengluGlyLe 46
 Db 179 AACCTTAGGAAGAACTCGAAGAGTGTGAAATTAAGAGGGGTCCGAGAACTACCTC 238
 QY 47 GluValValPheGlyLeuLeuLeuTyralaLeuAlaIleThrHisSerProLysVal 66
 Db 239 GGTCTGCGGTGAACCTTTGTTGTAATCTCTGCGACGATAGGACCTTCAATTAAGAT 298
 QY 67 GlnArgAlaAspSerAspValIlePheSerAsnSerPhe---GlyGluArgAsnValVal 85
 Db 299 AGTGTCCAAACCGACGCTACCTTCATCAAGCTTCGTTGGTGGTGGAGAGAAATTTTC 358
 QY 86 ValThrGluGlyAspLeuLysLysValLeuAspGlyCysAlaProLeuThrArgPheThr 105
 Db 359 CTCACCTACCGGTGAATGTAATCTCTGCGACGCTCAAGAACTTTGAGAGAAAGCT 418
 QY 106 AsnLysLeuArgThrPheGlyArgThrPheThrGluAlaIleValAspPheCysIleAla 125
 Db 419 AACCAAGTGGTGTGTTCTGCGCACTTTTCAGAGAGATCATATCTTCCGAGAGAA 478
 QY 126 TyrLysHisLysLeuProGlnLeuAsnAlaIleGluLeuGlyLeuProAlaGluAsp 145
 Db 479 TACCGAGGAAAGTGGCTCGATGCTAGAGCCACCGTACCGTCTACCTGTAAGAT 538
 QY 146 SerTyrLeuAlaIleAspPheLeuGlyThrCysProLysLeuSerGluLeuGlnGlnSer 165
 Db 539 CACTACTTACGCTGATTCATATCGACATCAACCAACTTACTAGCTTACCAACAGGT 598

QY 166 ArglyMetPheAlaSerMetTyrAlaLeuylThrGlyGlyValAlaAsnThrPro 185
 Db 599 COTTCGCTCTGGCCCGCAAAACCCACTCACAGAGTTCTCG---TTCGAATCACCA 655
 QY 186 ValSerAsnLeuArgGlnLeuGlyArg 194
 Db 656 GTAACGATTGAAACAGCTCGTCTCGT 682

RESULT 5

AB076539
 ID AB076539 standard; cDNA, 1500 BP.

AC AB076539;
 XX
 XX

DT 21-NOV-2002 (first entry)
 XX

DE C. albicans BAX-associated cDNA fragment SEQ ID 503.

XX Bax; Bax-resistance; cytostatic; fungicide; immunosuppressive; virucide;
 KW vasotropic; vaccine; gene therapy; proliferative disorder; cancer;
 KM apoptosis; fungal; yeast; infection; autoimmune disease; ischaemia;
 KW neurodegeneration; cell death; ss.

XX Candida albicans.

XX W0200264766-A2.

XX 22-AUG-2002.

XX 21-DEC-2001; 2001WO-EP015398.

XX 22-DEC-2000; 2000EP-00870318.

XX 04-JAN-2001; 2001EP-00870002.

XX 09-JAN-2001; 2001EP-00870003.

XX (JANC) JANSEN PHARM NV.

XX Contreras RH, Eberhardt I, Luyten WHML, Reekmans RJ;
 PT WPI; 2002-667002/71.
 DR P-PSDB; ABG93273.

XX New isolated nucleic acid representing a synthetic BAX-gene, useful as
 PT medicament for treating, preventing and/or alleviating yeast or fungal
 PT infections or proliferative disorders, or for preventing apoptosis in
 PT certain diseases.

XX Claim 36; Fig 2; 344pp; English.

XX This invention describes a novel nucleic acid representing a synthetic
 CC Bax gene. The Bax gene of the invention is useful for identifying Bax-
 CC resistant yeast or fungi, identifying, or obtaining and identifying
 CC Candida spp. sequences that are differentially expressed in a pathway
 CC eventually leading to programmed cell death or identifying inhibitors or
 CC inhibitor sequences of Bax-induced cell death. The products of the
 CC invention have cytostatic, fungicide, immunosuppressive, virucide and
 CC vasotropic activity and can be used in vaccines or for gene therapy. The
 CC isolated nucleic acids, polypeptides, pharmaceutical compositions,
 CC antisense molecules and antibodies are useful as medicaments or in
 CC preparing a medicament for treating, preventing and/or alleviating
 CC diseases associated with yeast or fungi or proliferative disorders, such
 CC as cancer, or for preventing apoptosis in certain diseases. The compounds
 CC as polypeptides, or the genetically modified organism are useful for
 CC preparing a medicament for modifying the endogenous flora of humans or
 CC other mammals. The vaccine is useful for immunising against yeast or
 CC fungal infections. Apoptosis-related diseases include autoimmune disease,
 CC ischaemia, diseases related with viral infections or neurodegenerations.
 CC This sequence represents a polynucleotide associated with the Bax gene
 CC described in the disclosure of the invention

XX Sequence 1500 BP; 442 A; 291 C; 309 G; 458 T; 0 U; 0 Other;
 XX
 XX
 XX

Alignment Scores:
 Pred. No.: 0.576 Length: 1500
 Score: 89.50 Matches: 53
 Percent Similarity: 43.60% Conservative: 39
 Best Local Similarity: 25.12% Mismatches: 76
 Query Match: 9.03% Indels: 43
 DB: 6 Gaps: 12

US-09-613-486-15 (1-198) x AB076539 (1-1500)

QY 3 LeuMetSerAspSerAsnLeuSerAsnLeuValIleThrAspAlaSerSer---LeuAsn 21
 Db 517 TTGATGCTCTCTTGGAATTTGGACCGAGTCTTCTGCTACTGCTCTACCTGCTTTTGAAG 576
 QY 22 GlyValAspIlyLysLeuLeuSerAla---GluValGluLysMetLeuValGlnLysGly 40
 Db 577 ACTGCTAAATCCACCCCATTTATCTGCTTTATCTTCCATTTGCTAGTCGAGCCGCT 636
 QY 41 AlaprobansnglylyllegluValAlapheglyLeuLeuLeuTyrAlaLeuAlaAlaArg 60
 Db 637 ATGCCAAGAAGGTATTCACACATTTGCTGTTTGTGCTACTGCTGCTGCTGCTCAT 696
 QY 61 ThrThrSerProIlyValGlnArgAlaAspSerAspValIlePheSerAsnSerPheGly 80
 Db 697 GCTAAACATCCCAAGATTGAAAAA-----GTGCTTCACTGCTTCTACTGCC 744
 QY 81 GluArgAsnValValAl-----ThrGlnGlyAspLeuLysVal-----94
 Db 745 ACCGTTAAATATTCAGAGAATTTGCTGCTGATCAACATGAAAAAGTTACTTTGAA 804
 QY 95 LeuAspGlyCysAlaProLeuThrArgPheThrAsnLysLeuArgThrPheGlyArgThr 114
 Db 805 TTGGGTGGTAAATCTCCCAACATTTGTTTCAAGAT-----GCTGATTTGGACAGACT 858
 QY 115 PheThrGluAlaTyrVal-----AspPheCysIleAla-----125
 Db 859 ATTCAAACTGATTTGTTTCTATCTTCTACAAATTCGTGAAGTCTGTGTGCTGTTCT 918
 QY 126 -----TyrIlyHisLysLeuProGlnLeuAsnAlaAla 137
 Db 919 COTCTTTGATTCATCCGCTGTGTACGACCACTGTTGAAAAATTCAGAAAGCTGCT 978
 QY 138 Glu-----LeuGlyIlePro---AlaGluAspSerTyrLeuAlaAlaAspPheLeu 153
 Db 979 GAAAGTGCAGAGTTGTTACCCATTCGACGAAACACTTTATAGGGTCC-----1029
 QY 154 GlyThrCysProIlyLeuSerGlnLeuGlnGlnSerArgLysMetPheAlaSerMetTyr 173
 Db 1030 -----CAAGTTCTGACGCTCAATTGTC---AAATTTGAAATACGTTGAA 1074
 QY 174 AlaLeuLysThrGlnGlyGlyValAlaAsnThr 184
 Db 1075 TCTGTAAATCTCAAGGTCTACTGTTGTTTAC 1107
 XX RESULT 6
 XX AAC45964 standard; DNA; 1197 BP.
 XX AAC45964;
 XX 18-OCT-2000 (first entry)
 XX Arabidopsis thaliana DNA fragment SEQ ID NO: 48418.
 XX Arabidopsis thaliana DNA fragment SEQ ID NO: 48418.
 XX Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway; metabolic pathway;
 KW promoter; termination sequence; ss.
 XX Arabidopsis thaliana.
 XX
 XX EP1033405-A2.
 XX
 XX 06-SEP-2000.
 XX
 XX

Fri Apr 9 15:30:05 2004

us-09-613-486-15.p2n.rng

```
XX 25-FEB-2000; 2000EP-00301439.
PF
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130044P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-013256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135622P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137503P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140911P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144006P.
PR 16-JUL-1999; 99US-0144085P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 25-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147316P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 15-AUG-1999; 99US-0149175P.
PR 17-AUG-1999; 99US-0149326P.
PR 18-AUG-1999; 99US-0149422P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-015138P.
PR 01-SEP-1999; 99US-015130P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
```


PR 24-SEP-1999; 99US-0155659P.
PR 26-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158022P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158365P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:
Pred. No.: 1.16 Length: 1197
Score: 86.00 Matches: 49
Percent Similarity: 38.92% Conservative: 23
Best Local Similarity: 26.49% Mismatches: 93
Query Match: 8.68% Indels: 20
DB: 3 Gaps: 9

US-09-613-486-15 (1-198) x AAC45964 (1-1197)

QY 7 SerAsnLeuSerAsnLeuValIleThrAspAlaSerSerLeuAsnGlyVal-----Asp 24
Db 427 TCGGAAACAACCAACATGTCACAGGTACATGCTGCTCAAGGATGTCGAGAC 486
QY 25 LysLysLeuLeuSerAlaGluValGluLysMetLeuValGlnLysGlyAlaProAsnGlu 44
Db 487 AAGAAAGCAGATGAGATCAATAACATGCCCTGCTGTGAGTCGAGGCTTAACGAT 546
QY 45 GlyIle-----GluValValPheGlyLeuLeuLeuValAlaLeuAlaArg 60
Db 547 TTCAATTTGAACTATTAAGAAATCCCTCAAGGGCTCCAGATTCCTTCATTCGTGT 606
QY 61 ThrThrSerProLysValGlnArgAlaAspSerAspValIlePheSerAsnSerPheGly 80
Db 607 TACCAAGACTTATCTTACAGAGGCTTGAAATTCGTCGCGAGCTTACAGTTAGT 666
QY 81 GluArgAsnValValValThrGluLysAspLeuLysValLeuAspGlyCysAlaPro 100
Db 667 GTCCGGAATGTTTGTGTT-----CGAGCTTACCGCAAG-----GGTGTTCACG 714
QY 101 LeuThrArgPheThrAsnLysLeuArgThrPheGlyArgThrPheThrGluAlaVal 120
Db 715 ATC--CATATGACTGCTAAATTCGCAACATTTTTCATTCCTTGGAAACACCATAC 771
QY 121 AspPheCysIleAlaValTyrLysIleLysLeuProGlnLeuAsnAlaAlaGluLeuGly 140

Db 772 AAAGACTCTGTTTATACATAGAGAACTCTTCGATCTTCCCAATGCAAGCATCT 831
QY 141 IleProAlaGluAspSerTyrLeuAlaAlaAspPheLeuGlyThrCysProLysLeuSer 160
Db 832 CTTCAGCA---AGCAAAATCTCTTACCCCAATGCTTATAT-----CTATGATGAG 862
QY 161 GluLeuGlnGlnSerArgLysMetPheAlaSerMetTyrAlaLeuLys---ThrGluGly 179
Db 883 ATGATCCAAACCCCTAGCAAA-----TACGGTTCAAGAGCAGCAAGAGA 927
QY 180 GlyValValAsnThr 184
Db 928 GGATGTTGTGAAACA 942
RESULT 7
AAC39012
ID AAC39012 standard; DNA; 1199 BP.
AC AAC39012;
XX 17-OCT-2000 (first entry)
DT
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 23034.
DE
XX Hybridization assay; genetic mapping; gene expression control;
KW Protein identification; signal transduction pathway; metabolic pathway;
KW Promoter; termination sequence; ss.
XX Arabidopsis thaliana.
OS
XX EP1033405-A2.
PN
XX 06-SEP-2000.
PD
XX 25-FEB-2000; 200DEP-00301439.
PF
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 29-MAR-1999; 99US-0126264P.
XX 01-APR-1999; 99US-0126785P.
XX 06-APR-1999; 99US-0127462P.
XX 08-APR-1999; 99US-0128234P.
XX 16-APR-1999; 99US-0128714P.
XX 19-APR-1999; 99US-0128845P.
XX 21-APR-1999; 99US-0130077P.
XX 23-APR-1999; 99US-0130510P.
XX 28-APR-1999; 99US-0130891P.
XX 30-APR-1999; 99US-0132048P.
XX 30-APR-1999; 99US-0132407P.
XX 04-MAY-1999; 99US-0132484P.
XX 05-MAY-1999; 99US-0132485P.
XX 06-MAY-1999; 99US-0132486P.
XX 06-MAY-1999; 99US-0132487P.
XX 07-MAY-1999; 99US-0132863P.
XX 11-MAY-1999; 99US-0134256P.
XX 14-MAY-1999; 99US-0134218P.
XX 14-MAY-1999; 99US-0134219P.
XX 14-MAY-1999; 99US-0134221P.
XX 14-MAY-1999; 99US-0134370P.
XX 18-MAY-1999; 99US-0134768P.
XX 19-MAY-1999; 99US-0134941P.
XX 20-MAY-1999; 99US-0135124P.
XX 21-MAY-1999; 99US-0135353P.
XX 24-MAY-1999; 99US-0135629P.
XX 25-MAY-1999; 99US-0136021P.
XX 27-MAY-1999; 99US-0136392P.
XX 28-MAY-1999; 99US-0136782P.
XX 01-JUN-1999; 99US-0137222P.

PR 03-JUN-1999; 99US-0137528P.
 PR 04-JUN-1999; 99US-0137562P.
 PR 07-JUN-1999; 99US-0137724P.
 PR 08-JUN-1999; 99US-0138094P.
 PR 10-JUN-1999; 99US-0138540P.
 PR 10-JUN-1999; 99US-0138847P.
 PR 14-JUN-1999; 99US-0139119P.
 PR 16-JUN-1999; 99US-0139452P.
 PR 16-JUN-1999; 99US-0139453P.
 PR 17-JUN-1999; 99US-0139452P.
 PR 18-JUN-1999; 99US-0139454P.
 PR 18-JUN-1999; 99US-0139455P.
 PR 18-JUN-1999; 99US-0139456P.
 PR 18-JUN-1999; 99US-0139457P.
 PR 18-JUN-1999; 99US-0139458P.
 PR 18-JUN-1999; 99US-0139459P.
 PR 18-JUN-1999; 99US-0139460P.
 PR 18-JUN-1999; 99US-0139461P.
 PR 18-JUN-1999; 99US-0139462P.
 PR 18-JUN-1999; 99US-0139750P.
 PR 18-JUN-1999; 99US-0139763P.
 PR 18-JUN-1999; 99US-0139817P.
 PR 21-JUN-1999; 99US-0139899P.
 PR 22-JUN-1999; 99US-0140353P.
 PR 23-JUN-1999; 99US-0140354P.
 PR 24-JUN-1999; 99US-0140695P.
 PR 28-JUN-1999; 99US-0140823P.
 PR 29-JUN-1999; 99US-0140991P.
 PR 30-JUN-1999; 99US-0141289P.
 PR 01-JUL-1999; 99US-0141842P.
 PR 01-JUL-1999; 99US-0142154P.
 PR 02-JUL-1999; 99US-0142055P.
 PR 06-JUL-1999; 99US-0142390P.
 PR 08-JUL-1999; 99US-0142803P.
 PR 09-JUL-1999; 99US-0142920P.
 PR 12-JUL-1999; 99US-0142972P.
 PR 13-JUL-1999; 99US-0145442P.
 PR 14-JUL-1999; 99US-0145624P.
 PR 15-JUL-1999; 99US-0144005P.
 PR 16-JUL-1999; 99US-0144086P.
 PR 16-JUL-1999; 99US-0144086P.
 PR 19-JUL-1999; 99US-0144325P.
 PR 19-JUL-1999; 99US-0144331P.
 PR 19-JUL-1999; 99US-0144332P.
 PR 19-JUL-1999; 99US-0144333P.
 PR 19-JUL-1999; 99US-0144334P.
 PR 19-JUL-1999; 99US-0144335P.
 PR 20-JUL-1999; 99US-0144352P.
 PR 20-JUL-1999; 99US-0144632P.
 PR 20-JUL-1999; 99US-0144884P.
 PR 21-JUL-1999; 99US-0144814P.
 PR 21-JUL-1999; 99US-0145086P.
 PR 21-JUL-1999; 99US-0145086P.
 PR 21-JUL-1999; 99US-0145088P.
 PR 22-JUL-1999; 99US-0145085P.
 PR 22-JUL-1999; 99US-0145087P.
 PR 22-JUL-1999; 99US-0145089P.
 PR 22-JUL-1999; 99US-0145192P.
 PR 23-JUL-1999; 99US-0145192P.
 PR 23-JUL-1999; 99US-0145192P.
 PR 23-JUL-1999; 99US-0145248P.
 PR 26-JUL-1999; 99US-0145276P.
 PR 27-JUL-1999; 99US-0145913P.
 PR 27-JUL-1999; 99US-0145918P.
 PR 27-JUL-1999; 99US-0145919P.
 PR 28-JUL-1999; 99US-0145919P.
 PR 28-JUL-1999; 99US-0146386P.
 PR 02-AUG-1999; 99US-0146388P.
 PR 02-AUG-1999; 99US-0146389P.
 PR 03-AUG-1999; 99US-0147038P.
 PR 04-AUG-1999; 99US-0147204P.
 PR 04-AUG-1999; 99US-0147302P.
 PR 05-AUG-1999; 99US-0147192P.

PR 05-AUG-1999; 99US-0147260P.
 PR 06-AUG-1999; 99US-0147303P.
 PR 06-AUG-1999; 99US-0147416P.
 PR 09-AUG-1999; 99US-0147493P.
 PR 09-AUG-1999; 99US-0147935P.
 PR 10-AUG-1999; 99US-0148177P.
 PR 11-AUG-1999; 99US-0148319P.
 PR 12-AUG-1999; 99US-0148341P.
 PR 13-AUG-1999; 99US-0148565P.
 PR 13-AUG-1999; 99US-0148684P.
 PR 16-AUG-1999; 99US-0149368P.
 PR 17-AUG-1999; 99US-0149175P.
 PR 18-AUG-1999; 99US-0149425P.
 PR 20-AUG-1999; 99US-0149723P.
 PR 20-AUG-1999; 99US-0149929P.
 PR 23-AUG-1999; 99US-0149902P.
 PR 25-AUG-1999; 99US-0150566P.
 PR 26-AUG-1999; 99US-0150894P.
 PR 27-AUG-1999; 99US-0151065P.
 PR 27-AUG-1999; 99US-0151066P.
 PR 27-AUG-1999; 99US-0151080P.
 PR 30-AUG-1999; 99US-0151303P.
 PR 31-AUG-1999; 99US-0151438P.
 PR 01-SEP-1999; 99US-0151930P.
 PR 07-SEP-1999; 99US-0152363P.
 PR 10-SEP-1999; 99US-0153070P.
 PR 15-SEP-1999; 99US-0153758P.
 PR 16-SEP-1999; 99US-0154018P.
 PR 20-SEP-1999; 99US-0154039P.
 PR 22-SEP-1999; 99US-0154779P.
 PR 22-SEP-1999; 99US-0155139P.
 PR 22-SEP-1999; 99US-0155486P.
 PR 24-SEP-1999; 99US-0156559P.
 PR 28-SEP-1999; 99US-0156458P.
 PR 29-SEP-1999; 99US-0156596P.
 PR 04-OCT-1999; 99US-0157117P.
 PR 05-OCT-1999; 99US-0157753P.
 PR 06-OCT-1999; 99US-0157655P.
 PR 07-OCT-1999; 99US-0158029P.
 PR 08-OCT-1999; 99US-0158232P.
 PR 12-OCT-1999; 99US-0158293P.
 PR 13-OCT-1999; 99US-0159293P.
 PR 13-OCT-1999; 99US-0159294P.
 PR 13-OCT-1999; 99US-0159329P.
 PR 14-OCT-1999; 99US-0159330P.
 PR 14-OCT-1999; 99US-0159331P.
 PR 14-OCT-1999; 99US-0159637P.
 PR 14-OCT-1999; 99US-0159638P.
 PR 18-OCT-1999; 99US-0159584P.
 PR 21-OCT-1999; 99US-0160741P.
 PR 21-OCT-1999; 99US-0160767P.
 PR 21-OCT-1999; 99US-0160768P.
 PR 21-OCT-1999; 99US-0160770P.
 PR 21-OCT-1999; 99US-0160814P.
 PR 21-OCT-1999; 99US-0160815P.
 PR 22-OCT-1999; 99US-0160980P.
 PR 22-OCT-1999; 99US-0160981P.
 PR 22-OCT-1999; 99US-0160989P.
 PR 25-OCT-1999; 99US-0161404P.
 PR 25-OCT-1999; 99US-0161405P.
 PR 25-OCT-1999; 99US-0161466P.
 PR 26-OCT-1999; 99US-0161359P.
 PR 26-OCT-1999; 99US-0161360P.
 PR 26-OCT-1999; 99US-0161361P.
 PR 28-OCT-1999; 99US-0161920P.
 PR 28-OCT-1999; 99US-0161922P.
 PR 28-OCT-1999; 99US-0161993P.
 PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:

Pred. No.:	1,16	Length:	119
Score:	86.00	Matches:	49
Percent Similarity:	38.92%	Conservative:	23
Best Local Similarity:	26.49%	Mismatches:	93
Query Match:	8.68%	Indels:	20
DB:	3	Gaps:	9

US-09-613-486-15 (1-198) X AAC39012 (1-1199)

Oy		, SerHisLeuSerLeuValIleThrAspAlaSerSerLeuAsnGlyVal-----Asp 24
Db	429	TCGGAAACACCAACAGTTTCCAAAGATTACATTGCTCTCTCCAAAGGATATGTGGAGAC 488
Oy	25	LyLeuSerLeuSerLeuValGluValGluGlyMetLeuValGlnGlyAlaProAsnGlu 44
Db	489	AAGAAACCGATGGAGATCATAAACATGCGCTTCGTGGTGAGGACAGGCCCTAAAGAT 548
Oy	45	GlyIle-----GluValValPheGlyLeuLeuLeuTyrAlaLeuAlaArg 60
Db	549	TTCAATTTCGACTATTACGAAATTCCTCCAGGCGCTCGAGTATCCTTCATTTCGCT 608
Oy	61	ThrThrSerProIysValGlnArgAlaAspSerAspValIlePheSerAsnSerPheGly 80
Db	609	TACCAAGACTTTATCTTTTAAAGAGCTTGAAATTTCTGTCGGAGACTTTAAAGTTTAG 668
Oy	81	GluArgAsnValValIleThrGluGlyAspLeuGlyIysValLeuAspGlyCysAlaPro 100
Db	669	GTCGGAGATGTTTTGGTT-----GGAGGTTTACCGCCATG-----GGGTGTTTACG 716
Oy	101	LeuThrArgPheThrAsnLysLeuArgThrPheGlyArgThrPheThrGluAlaTyrVal 120
Db	717	ATC---CATATGACTGCTGAATTCGCCCAACATTTTAGATTCTGCTGGAAACACCATPAC 773
Oy	121	AspPheCysIleAlaTyrLysHisLysLeuProGlnLeuAsnAlaAlaGluLeuGly 140
Db	774	AAAGACTCTGTTTATTCACATGAGAACTTCAGAACTCTTCCGCCCAATGAGAGACTCT 833
Oy	141	IleProAlaGluAspSerTyrLeuAlaAlaAspPheLeuGlyThrCysProLysLeuSer 160
Db	834	CTTCCAGAA---AGCAATATCTCTTTAGCCGAGTGTATAT-----CTATGATGGAG 884
Oy	161	GluLeuGlnGlnSerArgLysMetPheAlaSerMetTyrAlaLeuLys---ThrGluGly 172
Db	885	ATGATCCCAAAACCTTAGCAAA-----TACGGGTTCAAAAGACGCAAGAGA 922
Oy	180	GlyValValaLeuThr 184
Db	930	GGATGTTGTGGAACH 944

RESULT 8
AAAF86431
ID AAAF86431 standard; DNA; 349980 BP

AC	AAAF86431;
XX	
DT	29-OCT-2001 (first entry)
XX	
DE	Pyrococcus abyssi genomic fragment #1.
XX	
KM	Hyperthermophilic archaeon; hyperthermophilic protein; ds.
XX	
OS	Pyrococcus abyssi.
XX	
PH	Key
FT	Location/Qualifiers
FT	misc_feature
FT	300001..349980
FT	/*lag= a
FT	/note= "This sequence overlaps with the 5' end of
XX	AAH41223"
XX	
PN	FR2792651-A1.
XX	
PD	27-OCT-2000.
XX	

PF 21-APR-1999; 99FR-00005034.
 XX
 FR 21-APR-1999; 99FR-00005034.
 XX
 PA (CNRS) CNRS CENT NAT RECH SCI.
 PA (IFRE-) IFREMER INST FR RECH EXPL MER.
 XX
 XX Porterre P, Thiery JC, Prieur D, Dietrich J, Lecompte O;
 PI Querellou J, Weissenbach J, Saurin W, Helling R;
 XX WPI; 2001-126236/14.
 DR
 XX New nucleotide sequences isolated from *Pyrococcus abyssi* encode proteins
 PT useful in industry.
 PT
 XX
 XS Claim 1; Page 183-279; 1657p; French.

CC The present invention relates to the amino acid sequence of Pyrococcus
CC *abyssi* and P. *abyssi* proteins (see AB96053-AB966842). P. *abyssi* is a
CC hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal
CC vents. The present sequence is a fragment of the genomic sequence of P.
CC *abyssi*. The 3' end of this sequence overlaps with the 5' end of AA41223
CC *abyssi*.
CC The proteins of the present invention have various potential industrial
CC uses, since the proteins are stable at very high temperatures, some up to
CC 110 degrees centigrade. Note: This patent is in the same patent family as
CC WO00065062, which contains additional sequences as shown in AB99132-
CC AA99143, AAH75903-AAH75920 and AAG66436
XX
SQ Sequence 349980 BP; 95167 A; 76444 C; 78870 G; 97499 T; 0 U; 0 Other;

Alignment Scores:	
Pred. No.:	3.37e+03
Score:	86.00
Percent Similarity:	35.71%
Best Local Similarity:	25.97%
Query Match:	8.66%
D3:	5
Length:	39980
Matches:	40
Conservative:	15
Mismatches:	41
Indels:	5
Gaps:	8

US-09-613-486-15 (1-198) X AAF86431 (1-349980)

```

QY      26   LysLeuLeuSerAlaGluValGlnLysMetLeuValGlnIlySGLYALAProKengUGLY 45
Db      138056 AAAATTGGAGACCTTCATAGTGTAAAAAAGTGCTGAAGAATA---GTTCTCGGATTAAC 138112
QY      46   IleGluValValPheGlyLeuLeuLeuTyralaLeuAlalaArgThrThrserProlys 65
Db      138113 ATTACCATTAAT----- 138124
QY      66   ValGlnAlaGAlaAspSerAspValIlePheSerAsnserPheGlyGlnArgAsnVal---- 84
Db      138125 -----AAAGATGGGAGACGTGTGCATGCAAAAGGCGTTGGATACAGAAATGTTGAG 138175
QY      84 ----- 84
Db      138176 GCAAGCTTACATCTAACCCCGAGACAATATACGAATTGTTCAATAACCAAGACTTT 138235
QY      85 -----ValValThrGlnGlyAspLeuLysValLeuAspGly 97
Db      138236 ACAGCCTTGGCAATATAGAAATTAATGGAAGAGGGCGCATTAAGT-----TTGGATGAC 138289
QY      98   CysAlaProLeuThrArgPheThrAsn--LysLeuArgThrPheGlyArgThrPheThr 116
Db      138290 ----CCCGTTGAGAAATTCGTTAATATAAAGCTTAGACCATTGGAGAACACAGAGACC 138343
QY      117   GluAlaItyValAspPheCysIleAlaItyrLysHisLysLeuProGlnLeuAsnAlaIa 136
Db      138344 -----GTTTCATCACCTGTTAAGCAATTCCTCGGGAGATTCATCTTAGAGATTGCC 138399
QY      137   AlaGluLeu-----GlyIleProAlaGluAspSerTyrlen 148
Db      138395 GAGGCTTCATAGACGAATATGTCGGTGGGGAATACTGGTTG 138436

```


CC tissue origin, which may be used in numerous diagnostic and prognostic
 CC applications as well as cancer type-specific targets for therapeutic
 CC intervention. The compounds that modulate the activity of a tumour
 CC suppressor gene are useful in the treatment of cancer or as anti-cancer
 CC drugs. The present sequence represents a polynucleotide of the invention.

XX Sequence 5311 BP; 1462 A; 1175 C; 1235 G; 1439 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	13.1	53.11	42	28	63	50
Percent Similarity:	85.00					
Best Local Similarity:	38.25%					
Query Match:	22.95%					
DB:	9					9

US-09-613-486-15 (1-198) x ADD29639 (1-5311)

QY 8 AsnLeuSerHisLeuValIleThrAspAlaSerSerLeuHisnGlyValAspIleuSleu 27
 DB 2030 AATATCCAAATATATCTT-----GCTCAAAACCTAAGTGGGGACAGAAAGAAAG 2080
 QY 28 LeuSerAlaGluVal-----GluIleMetLeuValGlnIleGlyAlaProAsn 43
 DB 2081 CTGACTTTGGAAATCGCCATTAGAGATTCAGATTCTTCTAGCTGATGAGCCACC 2140
 QY 44 GluGlyIleGluVal-----ValPheGlyLeuLeuLeuTyralaLeuAla 58
 DB 2141 GCTGCTGCTGATCCCTTTTCAAGGACCGGCTGGGATCTCTG----- 2185
 QY 59 AlaArgThrThrSerProIleValGlnArgAlaAspSerAspValIlePheSerAsnSer 78
 DB 2186 -----AAGAGCGCAGAGCAGACCGCTGGTCTCTG----- 2185
 QY 79 Phe-----GlyGluArgAsnValValValThrGluGlyAspLeu 91
 DB 2228 TTCAATGATGAGGCTGACATCTGCTGATAGGAAAGTTCATCTCCATGGAGAGCTG 2287
 QY 92 LysIleValLeuAspGlyCysAlaPheLeuThrArgPheThrAsnIleuArgThrPhe 111
 DB 2288 AAG-----TGTCAGGATCATCTCTTCTTCAAGAAAGAAATGGGAGCTT 2332
 QY 112 GlyArgThrPheThrGluAlaIleValAspPheCysIle-----Ala 125
 DB 2333 GGGATACCTTAAGTTGACAGCTGAAGAAGCTGTCTTCAGAAACATCAGATGCTT 2392
 QY 126 TyrIleHisIleuPheGlnIleuAsnAlaAlaIleu-----Leu 139
 DB 2393 GTTAAACAGACATCCCTGACGAGCAAGCTATCAGCTTAAGGGGAGAAAGCTTCTAT 2452
 QY 140 GlyIleProIleGluAspSerTyrLeuAlaAlaIleu-----LeuGlyThrCys 156
 DB 2453 ACATACCTTAAAGAAACATTAATATTCAGAGCTTGCAGAGCTTGCAGAGCTGT 2512
 QY 157 ProIleu 159
 DB 2513 CCTGAGCTG 2521

RESULT 11

AAA91259
 ID AAA91259 standard; DNA; 4766 BP.

AAA91259;

06-AUG-2003 (revised)

08-MAY-2001 (first entry)

GLRAV-5 nucleotide sequence.

GLRAV-5; grapevine leafroll virus; GLRAV infection; GLRAV coat protein;

GLRAV HSP70 homologue protein; viral gene mapping;

plant disease resistance; ds.

OS Grapevine leafroll virus.

XX Key Location/Qualifiers

PH Key 1..1593

FT CDS /tag= a

FT /product= "HSP70 homologue protein"

FT /partial /note= "No start codon given, Specifically claimed region"

FT CDS /tag= b

FT /product= "ORF 2 protein"

FT /note= "Specifically claimed region"

FT CDS /tag= c

FT /product= "coat protein"

FT /note= "Specifically claimed region"

FT CDS /tag= d

FT /product= "duplicate coat protein"

FT /note= "Specifically claimed region"

PA WO200105957-A2.

XX 25-JAN-2001.

XX 19-JUL-2000; 2000MO-US019708.

XX 19-JUL-1999; 99US-0144453P.

XX (AGRI-) AGRITOP INC.

XX Good XC, Monis J;

XX WPI, 2001-147339/15.

XX P-PSDB; AAY97688, AAY97689, AAY97690, AAY97691.

XX Novel grapevine leafroll virus polynucleotide useful as diagnostic and

XX probe, for viral gene mapping and for induced plant disease resistance.

XX Claim 1; Fig 1; 60pp; English.

XX This sequence represents a grapevine leafroll virus (GLRAV-5) DNA

XX sequence of the invention. The DNA sequence can be used in an expression

XX construct. The construct is useful for providing resistance to GLRAV

XX infection in a recombinant plant cell by transforming the plant cell with

XX it, where transcription of the polynucleotide sequence interferes with a

XX normal viral function such as movement, encapsidation or replication of

XX viral RNA. The polynucleotide sequence is expressed as an antisense

XX sequence and encodes a GLRAV coat protein, preferably a defective GLRAV

XX coat protein or a GLRAV HSP70 homologue protein. The GLRAV-5 DNA is

XX useful for the synthesis of GLRAV, as diagnostics and probes, for viral

XX gene mapping and for induced plant disease resistance. It is also useful

XX to detect and quantitate expression of GLRAV in plant tissue prior to use

XX in vegetative propagation, by detecting the presence of GLRAV RNA.

XX (Updated on 06-AUG-2003 to correct OS field.)

XX Sequence 4766 BP; 1329 A; 890 C; 1231 G; 1316 T; 0 U; 0 Other;

XX Alignment Scores:

XX Pred. No.:

XX Score:

XX Percent Similarity:

XX Best Local Similarity:

XX Query Match:

XX DB:

XX US-09-613-486-15 (1-198) x AAA91259 (1-4766)

QY 61 ThrThrSerProIleValGlnArgAlaAspSerAspValIlePheSerAsnSerPheGly 80

DB 3663 TCACGCTCTCTTAAGGTGTCACGCTTCTTCATCGAGCTATA--ACCGCTAAGTATGAT 3719

QY 81 GluArgAsnValIValThrGluGlyAspLeuLysValLeuAspGlyCysAlaPro 100
 Db 3720 GGTAAAGAGTACAGTGGCCCATGATGATTAAGACAGCTTAGCAATCTATAGGG 3779
 QY 101 LeuThrArgPheThrAsnLysLeuAlaGthrPheGlyArgThrPheThrGluAlaVal 120
 Db 3780 TCTTCGGTTACGAAAACCTCTTAGCAATTTGAGAGCATTCAGCGCATATGTC 3839
 QY 121 AspPheCysIleAlaTyLysHisLysLeuPProGlnLeuAsnAlaAlaGluLeuGly 140
 Db 3840 CAAGGATATCTCTGAGCAAACTGGAGGTG---AAGACGAAGATCGGCTTTCACGGT 3896
 QY 141 IleProAlaGlu----- 144
 Db 3897 GTGCGCGCTAACTACTACTCTGCGACAGCTGCTGCATGTTGATGCAAGCTGTC 3956
 QY 145 -----AspSerTyLysAlaAlaAspPheLeuGlyThrCysProLysLeuSerGluLeu 162
 Db 3957 GGTATGACGCTTCACTGCGAGCTGAG---TTAGGCAAGATGTTGTCATCAATAAACT 4013
 QY 163 GlnGlnSerArgLysMetPheAlaSerMetCys 173
 Db 4014 TCCATAGCAACAGAGCTACACACATCTGTAC 4046

RESULT 12
 ACN29885
 ID ACN29885 standard; DNA, 1266 BP.
 XX
 ACN29885;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Prokaryotic essential gene #11542.
 XX
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;
 XX
 OS Corynebacterium diptheriae.
 XX
 PN WC200277183-42.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 XX
 PR 06-SEP-2001; 2001US-00948983.
 XX
 PR 25-OCT-2001; 2001US-0342922P.
 XX
 PR 08-FEB-2002; 2002US-00072851.
 XX
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen XL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 DR WPI; 2003-029926/02.
 XX
 PS P-PSDB; ABU26015.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 Claim 14; SEQ ID NO 17755; 1766pp; English.

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
 CC prokaryotic essential genes. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX

SO Sequence 1266 BP; 330 A; 280 C; 321 G; 335 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 2.6 Length: 1266
 Score: 83.50 Matches: 46
 Percent Similarity: 32.62% Conservative: 15
 Best Local Similarity: 24.60% Mismatches: 67
 Query Match: 8.43% Indels: 59
 DB: 7 Gaps: 7

US-09-613-486-15 (1-198) x ACN29885 (1-1266)

QY 53 LeuLeuTyRAlaLeuAlaAlaArgThrSerProLysValGlnArgAlaSerSer 72
 Db 316 TTRATTCGACGCTTCGCGAGAGAAACCAACGACTCAACATCTCACCATCTCGA 375
 QY 73 ValIlePheSerAsnSerPheGlyGluArgAsnValIValThrGluGlyAspLeu 92
 Db 376 GTATTTTCTGTACTCAGTACTAACTAAT----- 408
 QY 93 LysValLeuAspGlyCysAlaProLeuThrArgPheThrAsnLysLeuArgThrPhe 111
 Db 409 -----GAGCGGCATTCAAGTTGGCTGCTATGACTGGAAGAACCGAATATTAGCA 459
 QY 112 -----GlyArgThrPhe----- 115
 Db 460 GCGCAGCATGTTTTCACGCTGTCGATGAGACACTTCTATACGGGCCAAGCTGAC 519
 QY 116 -----ThrGluAlaTyRValAspPheCysIle 124
 Db 520 AACAATGCGGTTTATCCGCTGCGCCGAGCGTTGAATTTTATCCATGCGACATT 579
 QY 125 AlaTyLysHisLysLeuPProGlnLeuAsnAlaAla----- 136
 Db 580 GACTATTTGCGCAGGTTTTCACAACTCGACAGATGTCGACGCCATTAATTTTAGAG 639
 QY 137 -----AlaGluLeuGly-----IleProAlaGluAspSerTyRLeuAlaAspPhe 152
 Db 640 CCAATTCAGAGTGAACCTGTTGTTATTCCTGCTCCAGATGATTTTGTCTCG---GTT 696
 QY 153 LeuGlyThrCysProLys-----LeuSerGluLeuGlnLeuSerArg 166
 Db 697 AGGGGATTAATGCAATTAATTTGACATTCGATGATGTCGATGATGATGATGATGATG 756
 QY 167 LysMetPheAlaSerMetTyRAlaLeuLysThrGluGlyValValAlaSerThrProVal 186
 Db 757 GAGCGACAGTGAAGATTTTTCGATTTTCAGATGAAACGAGTGATATCCGACGCGTA 816
 QY 187 SerAsnLeuArgGlnLeuGly 193

Db 817 ACTATGCGCAAAAGTTTGGGA 837

RESULT 13

ABA92787.4/c

Continuation (5 of 7) of ABA92787 from base 400001 (Buchnera sp. genomic DNA seq ID NO:1
WP Sequence Split into 7 fragments LOCUS ABA92787 Accession ABA92787

Fragment Name	Begin	End
WP ABA92787_0	1	110000
WP ABA92787_1	100001	210000
WP ABA92787_2	200001	310000
WP ABA92787_3	300001	410000
WP ABA92787_4	400001	510000
WP ABA92787_5	500001	610000
WP ABA92787_6	600001	640681

Alignment Scores:

Pred. No.:	1.57e+03	Length:	110000
Score:	83.50	Matches:	32
Percent Similarity:	37.18%	Conservative:	26
Best Local Similarity:	20.51%	Mismatches:	53
Query Match:	8.43%	Indels:	45
DB:	6	Gaps:	5

US-09-613-486-15 (1-198) x ABA92787_4 (1-110000)

QY 65 LysValGlnArgAlaAspSerAspValIlePheSerAsnSerPheGlyGluArgAsnVal 84
Db 35915 AAAAATGACCTGGATGATTAAGCATGATGACGAGGCTCCAGGCTCAAAAGAAATAT 35856
QY 85 ValValThrGluGlyAspLeuLysLysValLeuAsp----- 96
Db 35855 TTATTTGAAATGGAATATTAAATAATATATGCAATTAATGCGGCTTTAATG 35796
QY 97 -----GlyCysAlaPheLeuThrArg 103
Db 35795 GGTGTAACTCTACCGAAATGTCGTGCGTGAATCTTATTCGTCTAATGCTGCTCGT 35736
QY 104 PheThrAsnLysLeuArgThrPheGlyArgThrPheThrGluAlaTyrValAspPheCys 123
Db 35735 ATGACTTAAT-----ACCTAATG 35718
QY 124 IleAlaTyrLysHisLysLeuProGlnLeuAsnAlaIleAlaGluLeuGlyIleProAla 143
Db 35717 TTATCTGGAAGTGAATATGATGATATATTAAGCCGCTGACTATGGAATA----- 35664
QY 144 GluAspSerLysLeuAlaAlaAspPheLeuGlyThrCysProLysLeuSerGluLeuGln 163
Db 35663 -----TGTGCTGTAATTTTTCGGA-----GCACGAGTAGATTAATCT 35625
QY 164 GlnSerArgLysMetPheAlaSerMetTyrAlaLeuLysThrGluGlyValValAsn 183
Db 35624 TCCGGAATAATTGTAATTTCTACTTCGGAAGCTTATTAATAAAAGTGAATAATGTT 35565
QY 184 ThrProValSerAsnLeuArgGlnLeuGlyArg---ArgGluValMet 198
Db 35564 ACACCAATTAAACACTACTCTCATAGATCAGGATTAAGAAATTAG 35517

RESULT 14

ACA47601
ID ACA47601 standard; DNA; 1773 BP.

XX AC ACA47601;
XX AC ACA47601;
XX DT 19-JUN-2003 (first entry)
XX DE Prokaryotic essential gene #29258.
XX KM Antisense; ds, prokaryotic essential gene; cell proliferation;
XX KM drug design; gene.
XX XX Staphylococcus haemolyticus.
XX

PM WO20027183-A2.

PD 03-OCT-2002.

PF 21-MAR-2002; 2002WO-US009107.

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-03429232.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

PA (ELITR) ELITRA PHARM INC.

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

PI P-PsDB; ABU43731.

PS Claim 14; SEQ ID NO 35471; 1766bp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

CC of the 6213 antisense sequences given in the specification where expression

CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

CC polypeptide or its fragment whose expression is inhibited by the

CC antisense nucleic acid; (4) an antibody capable of specifically binding

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

CC proliferation or the activity of a gene in an operon required for

CC the gene product or that has an activity against a biological pathway or

CC required for proliferation, or that inhibits cellular proliferation; (8)

CC identifying a gene required for cellular proliferation or the biological

CC pathway in which a proliferation-required gene or its gene product lies

CC or a gene on which the test compound that inhibits proliferation of an

CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene

CC product is overexpressed or underexpressed; (12) determining the extent

CC to which each of the strains is present in a culture or collection of

CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for

CC identifying proteins or screening for homologous nucleic acids required

CC for cellular proliferation to isolate candidate molecules for rational

CC drug discovery programs, or for screening homologous nucleic acids

CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pat_sequences

SQ Sequence 1773 BP; 602 A; 305 C; 335 G; 531 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	5.64	Length:	1773
Score:	82.50	Matches:	35
Percent Similarity:	37.88%	Conservative:	15
Best Local Similarity:	26.52%	Mismatches:	43
Query Match:	8.32%	Indels:	39
DB:	7	Gaps:	6

US-09-613-486-15 (1-198) x ACA47601 (1-1773)

QY 4 MetSerAspSerAsnLeuSerAsnLeuValIleThrAspAlaSerSerLeu----- 20
Db 865 ATGATGCTCTCTATGCAAGATATATGCTTAAACCGAATGATTTATTAATTA 924

```

QY 21 ---AsnGlyValAlaAspIysIleuSerAlaGluValGluLysMetLeuValGlnLys 39
DB 925 GGTAGTCGGTTTGTATGATTAAGCAAGT----- 954
QY 40 GlyAlaProAsnGluGlyIleGluValAlaPheGlyLeuLeuLeuTyrAlaLeuAla 59
DB 955 ---AATCCTATATGA-----TTTGCTCCTCATGCA 981
QY 60 ArgThrThrseryProIysValGlnArgAlaAspSerAspValIlePheSerAsnSerPhe 79
DB 982 ACCGTAGTCATGATGATATGATGATGAATTAACAGGTGATTCGACTGACTTA 1041
QY 80 GlyGluArgAsnValValValThrGluGlyAspLeuLysValLeuAspGlyCysAla 99
DB 1042 GGG-----ATAGTTCGCTGACTGTAAATAAGTACTCGAAGCTTATCT 1083
QY 100 ProLeuThrArgPheThrAsnLysLeuArgThrPheGlyArgThrPheThrGluAlaTyr 115
DB 1084 -----CAAGATTTCACAAATAAA-----ACTTCCCATGAAACACTGG 1119
QY 120 ValAspPheCysIleAlaTyrLysHisLysLeuPro 131
DB 1120 GTTAACTTGTATTGAAATTAAGCACAAACATCCT 1155

RESULT 15
AAA72313
ID AAA72313 standard; DNA; 1209 BP.
XX
XX AAA72313;
XX
XX 11-DEC-2000 (first entry)
XX
XX Erysipelothrix rhusiopathiae PPA fragment DNA.
XX
XX DE Protective polypeptide antigen; PPA; Fujisawa strain; swine erysipelas;
XX KM recombinant production; immunogenic; subunit vaccine;
XX KM transmembrane administration; ds.
XX
XX Erysipelothrix rhusiopathiae.
XX
XX OS
XX
XX Key Location/Qualifiers
XX FH 1. 1209
XX FT /tag= a
XX FT /product= "Erysipelothrix rhusiopathiae 46.5 kD PPA
XX FT fragment"
XX FT /note= "No initiation or termination codon given in the
XX FT specification"
XX
XX JP3072345-B1.
XX
XX 31-UTL-2000.
XX
XX 31-MAR-1999; 99JP-00094004.
XX
XX 31-MAR-1999; 99JP-00094004.
XX
XX (NORO) NORINSUSANSHO KACHIKU EISEI SHIKENJOCHO.
XX PA (HGET) HIGETA SHOYU KK.
XX PA (FUOI-) FUJITA GAKUEN.
XX
XX WPI; 2000-551432/51.
XX DR P-PSDB; AAB14800.
XX
XX Novel recombinant protective polypeptide antigen useful as a vaccine for
XX PT protecting livestock against Erysipelothrix rhusiopathiae infection.
XX
XX Claim 3; Fig 4A-4C; 23p; Japanese.
XX
XX This sequence represents DNA encoding a 46.5 kD immunogenic fragment of
XX CC the protective polypeptide antigen (PPA) of the bacterium Erysipelothrix
XX CC rhusiopathiae. This bacterium infects livestock, particularly pigs (swine
XX CC erysipelas), and is also able to cause disease in humans. The

```

```

CC recombinantly produced PPA fragment may be used as a subunit vaccine for
CC stimulating an immune response in animals against Erysipelothrix
CC rhusiopathiae. As the vaccine can be produced recombinantly, it can be
CC produced on a large scale. The vaccine is suitable for transmembrane
CC administration
XX
SQ Sequence 1209 BP; 456 A; 177 C; 237 G; 339 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 4.36 Length: 1209
Score: 81.50 Matches: 33
Percent Similarity: 40.60% Conservative: 21
Best Local Similarity: 24.81% Mismatches: 50
Query Match: 9.22% Indels: 29
DB: Gaps: 5
US-09-613-486-15 (1-198) x AAA72313 (1-1209)
QY 60 ArgThrThrseryProIysValGlnArgAlaAspSerAspValIlePheSerAsnSerPhe 79
DB 358 AGATTAAAGATTCTCGAATTCGATGAAGCAATACCTGATTTACTCGATTAAGAGGTG 417
QY 80 GlyGluArgAsnValValValThrGluGlyAspLeuLysValLeuAspGlyCysAla 99
DB 418 AAACACCGTATTATTAGTAAATATGAAGTAAAGTTAA-----GGTAAAGCT 465
QY 100 ProLeuThrArgPheThrAsnLysLeuArgThrPheGlyArgThrPheThr----- 116
DB 466 CCTTAAAGCATTTATAGTCTCTTAAGAGATGAATTCGATTAAGAAATTCCT 525
QY 117 -----GluAlaTyrValAspPheCysIleAla----- 125
DB 526 GCAGAAAGTAATTTATTTACCTGAAGCCGATGAGATTTCTTAGTTTCAGATTCAAGCGAG 585
QY 126 TyrLysHisLysLeuProGlnLeuAsnAlaAlaGluLysIleProAlaGluAsp 145
DB 586 TATTAATGACAACTAAATATATCAACTTCTTGGGTCTAGGGGTGAGGAGTTTAT 645
QY 146 SerTyrLeuAlaAlaAspPheLeuGlyThrCysProIysLeuSerGluLeuGlnGlnSer 165
DB 646 GACTAT-----AACCGGCTCGAAATATGATGAGAAAAA 678
QY 166 -----ArgLysMetPheAlaSerMetTyrAlaLeuLys 176
DB 679 GAACCTCATCCACTGTATCTTGAACTTATATGCTATGCGG 717

Search completed: April 7, 2004, 02:20:39
Job time : 394 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: April 7, 2004, 02:07:34 ; Search time 2353 Seconds
(without alignments)
2506.448 Million cell updates/sec

Title: US-09-613-486-15
Perfect score: 991
Sequence: 1 MEIMSDSNLSNLTVDASTL.....GGVNTPVSNLQGRREV 198

Scoring table:
BLOSUM62
Xgapod 10.0, Xgapext 0.5
Ygapod 10.0, Ygapext 0.5
Fgapod 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODBL=frame+g2n.model -DEV=x1h
-Q=/cgn2_1/USPTO.spool/US09613486/runat_05042004_155121_28274/app_query.fasta_1.391
-DB=EST -QMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOFCU=0 -LOOPEXT=0
-UNITS=bites -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=spc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09613486 @CNC 1.1 2135 @runat_05042004_155121_28274 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEOUTERY -NES_SCORES=0 -WAIT -DSPELCK=100 -LONGLOC
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREDS=1 -XGAPOD=10 -XGAPEXT=0.5 -FGAPOD=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estm:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_p1n:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vr1:*
28: gb_gss1.*

29: gb_gss2.*
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88.5	8.9	706	13 BQ986943	BQ986943 QGF10N21.
2	88.5	8.9	748	28 B2051926	B2051926 jms7h07.
3	87	8.8	477	12 BM132093	BM132093 TGEStx0b
4	87	8.8	539	14 CF267035	CF267035 TGEStx02
5	85.5	8.6	562	13 BQ459495	BQ459495 H408017r
6	85.5	8.6	666	13 BU0007723	BU0007723 QGHS515.Y
7	85.5	8.6	671	12 BJ433495	BJ433495 B433495
8	85.5	8.6	704	28 B2434184	B2434184 BONFR91TF
9	85.5	8.6	713	9 AU034876	AU034876 AU034876
10	85.5	8.6	751	12 BJ376355	BJ376355 BJ376355
11	85	8.6	654	13 BQ916284	BQ916284 QH17G04.
12	84.5	8.5	610	13 BQ846646	BQ846646 QGA7N01.Y
13	84.5	8.5	611	13 BQ984758	BQ984758 QGSAU05.Y
14	84.5	8.5	632	12 BG526441	BG526441 61-95 Ste
15	84.5	8.5	648	12 BG523153	BG523153 29-48 Ste
16	84.5	8.5	695	13 BQ993014	BQ993014 QGF27F14.
17	84	8.5	589	13 BQ983386	BQ983386 QGB16R23.
18	84	8.5	600	10 BF598403	BF598403 srt17b06.Y
19	84	8.5	732	28 BH976504	BH976504 odt67801.
20	84	8.5	1098	12 BG247808	BG247808 sc035915
21	83.5	8.4	433	14 CA169078	CA169078 SC0FSB101
22	83.5	8.4	633	28 BH096256	BH096256 RPCT-24-2
23	83.5	8.4	644	12 BJ340622	BJ340622 BJ340622
24	83.5	8.4	644	12 BJ343395	BJ343395 BJ343395
25	83.5	8.4	704	13 C921233	C921233 C921233 Dict
26	83.5	8.4	600	14 CB341251	CB341251 CA12EN000
27	83	8.4	755	29 CC722225	CC722225 OGVA288TH
28	83	8.4	911	12 BG421690	BG421690 602449789
29	82.5	8.3	661	28 AQ949879	AQ949879 Sheared D
30	82.5	8.3	762	14 CB569015	CB569015 AGENCOURT
31	82	8.3	607	12 BT775956	BT775956 f61h03.Y
32	82	8.3	627	13 BU026578	BU026578 OHG17R05
33	82	8.3	750	13 BQ481813	BQ481813 PV.GE8012
34	82	8.3	828	13 BQ410316	BQ410316 60515838
35	82	8.3	923	14 CA582924	CA582924 EST002599
36	82	8.3	1028	13 BU202935	BU202935 603102659
37	81.5	8.2	428	12 BG544734	BG544734 E2700 Ch1
38	81.5	8.2	560	28 CC155340	CC155340 CSU-K34.1
39	81.5	8.2	606	28 AZ079289	AZ079289 RPCT-23-4
40	81	8.2	513	28 AQ951071	AQ951071 Sheared D
41	81	8.2	727	28 BH929686	BH929686 odt98612.
42	81	8.2	807	28 B2577160	B2577160 msh2_5276
43	80.5	8.1	738	29 CG406315	CG406315 ZMWBB026
44	80.5	8.1	836	12 BI085103	BI085103 602870643
45	80	8.1	432	13 BU765335	BU765335 ssa816f01.

ALIGNMENTS

RESULT 1
BQ986943
LOCUS
DEFINITION BQ986943 706 bp mRNA linear EST 21-AUG-2002
ACCESSION QGF10N21, mRNA sequence.
VERSION BQ986943.1 GI:22404468
KEYWORDS EST.
SOURCE Lactuca sativa
ORGANISM Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; campanulids; Asterales; Asteraceae; Cichorioideae;

Chloritase; Lactuca.
1 (bases 1 to 706)
Kozik, R., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
Lin, H., van Damme, W., Lavelle, D., Chevalier, P., Ziegler, J.,
Ellison, P., Kolman, J., Slabaugh, N.S., Livingston, K., Zhou, Y.,
Lai, Z., Church, S., Jackson, L. and Bradford, K.
Lettuce and Sunflower ESTs from the Compositae Genome Project
http://compositae.ucdavis.edu/
Unpublished (2002)

JOURNAL
COMMENT
Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundo Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.org [michelmore@ucdavis.edu]
belongs to contig QC_Ca_contig3108, see http://cgpdb.ucdavis.edu/
for details.
Plate: QGF10 row: N column: 21.

FEATURES
source
Location/Qualifiers
1..706
/organism="Lactuca sativa"
/mol_type="mRNA"
/cultiVar="L.serriola"
/db_xref="taxon:4236"
/clone="QGF10N21"
/lab_host="E.coli"
/clone_lib="QG_EFGHJ lettuce serriola"
/note="Vector: pKODXAF1AB; The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at http://cgpdb.ucdavis.edu/
TAG_LIB=QG_EFGHJ lettuce serriola
TAG_SEQ=TCGTGCGG3"

ORIGIN

Alignment Scores:

Pred. No.: 1.45 Length: 706
Score: 88.50 Matches: 51
Percent Similarity: 41.36% Conservative: 28
Best Local Similarity: 26.70% Mismatches: 53
Query Match: 8.93% Indels: 13
DB: 13 Gaps: 59

US-09-613-486-15 (1-198) x BQ86943 (1-706)

QY 13 VALLIETTRASPALSERSEIUENGLYVALAPLYSLVLEULESERALAGLVAL 32
DB 137 GTTTGACCGAT---GGGTCTTACCATGATGTTGCTCGAGTTCCTTCAG--- 190
QY 33 GULYSMELEUVALGMLVGLYVALAProAENGUJLYIEGLUVALVALPhedLYLeu 52
DB 191 -----CTGCGACGACGAGGCGGCTTCGGAAGGCTT----- 223
QY 53 LEULEUTYRALALEUHLAALARThrThrSerProLYs-----ValGILPARG 68
DB 224 -----AGGAAAGCTGCTCCAAAGATCTTGAGCCGATTTATGAGA 262
QY 69 ALAASP-----SerASPVALLEPheser----- 76
DB 263 GTTGAAGTTGTGACCGCTTGAAGAACATTGGGGGATGTCATTGGGGATTGGAATTCGAGA 322
QY 77 -----AsnSerPhedLYGLUARGAsnVALVALValThrGLUGLYAspLeuLYs 92
DB 323 AGAGGCGACATTAATGATTTGTGTGATTAAG-----CCTGTGACACTT--- 364
QY 93 LYSVALLEUASPGLYCYVALAPROLEUTHr-----ArgPhdThrAsnLYsLeuARG 109

DB 365 AAGGTGTGATTCCTTGTGCTCCACTGCGGAATGCTTACGATGATGATTAAG 424
QY 110 -----ThrPhedLYArg-----ThrPhthrGLUAlaYrVALASPPhCYsILAlaLYr 126
DB 425 GGAATGACAAAGAGAGAGCTTCTTACCAATGCAATTAATTAATGATGCTTGCCT 484
QY 127 LYSLSLYLEUPROGLNLEUASNAALAAIAGLUEUGLYIEPROALIGLUSPser 146
DB 485 CAACATATTCAGAACACAGCTTCTGCTGTAAAGAAAGAGCTGCGACGTTGATCTTCA 544
QY 147 -----TYLEUAla-----ALAASPPhLEUGLYThrCYsPRO 157
DB 545 TCTTCATCTTGTGCTATTATTAATGATGAGGGAATTTCTGATTTCACCTCATATGCAA 604
QY 158 LYSLEUSERGLULEUGLINDSERARGLYSmet 168
DB 605 TATAGCAAAACAAAAACAAACAAAGGATT 637

RESULT 2
BZ051926/c 748 bp DNA linear GSS 09-OCT-2002
LOCUS jmr57h07.g1 B.oleracea001 Brassica oleracea genomic, genomic survey
DEFINITION sequence.
ACCESSION BZ051926 GI:23652910
VERSION BZ051926.1 GI:23652910
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 748)
Deleantony, K., Fewell, G., Fulton, L., McCombie, W.R., Miner, T.,
Nash, W., Rablowicz, P.D., and Wilson, R.K.
Whole genome shotgun reads from Brassica oleracea
Unpublished (2002)
JOURNAL
TITLE
REFERENCE
AUTHORS
COMMENT
Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: rkwilson@wustl.edu
Plate: jmr57 row: h column: 07
Seq primer: -28RPOT reverse
Class: shotgun
High quality sequence start: 56
High quality sequence stop: 551.
Location/Qualifiers
1..748
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.Oleracea001"
/note="Vector: pOTW13; Whole genome shotgun library from
flowering buds. DNA was purified from a crude nuclear
prep using Brassica oleracea TO1000DH3 buds provided by
Thomas Osborn at the University of Wisconsin. Genomic
DNA was provided by Pablo Rablowicz (GSH) and the
shotgun library prepared at Washington University genome
Sequencing Center."

FEATURES
source

ORIGIN

Alignment Scores:

Pred. No.: 1.59 Length: 748
Score: 88.50 Matches: 33
Percent Similarity: 47.00% Conservative: 14
Best Local Similarity: 33.00% Mismatches: 40
Query Match: 8.93% Indels: 13
DB: 28 Gaps: 5

US-09-613-486-15 (1-198) x BZ051926 (1-748)

QY 84 VALVALVALThrGLUGLYAspLeuLYsVALLeu-----AspGLYCYsALAPROLeu 101
::: ::::

Db 551 ATACCATCGTGAAGACCTACGACAAATCTTAGGAGTGTGAGTGCATGCTC 492
 Qy 102 ThrArgPheThrAsnLysLeuAArgThrPheGlyAArgThrPheThrGluAlaValAsp 121
 Db 491 ACTTTTAAGACTTAACAAGACGACATATAGGCCGATAGCTGTCAAGGTGACACAC 432
 Qy 122 PheCysIleAlaTyrLys--HisLysLeuProGlnLeuAsnAlaAlaGluLeuGly 140
 Db 431 TTGATGCTCAAGAGGAGGAGCTCAAGATTAACAGGGTTCATGACGAGCTGGCGGACT 372
 Qy 141 ILleProAlaGluAspSerTyrLeuAlaAlaAspPheLeuGlyThrCysProLysLeuSer 160
 Db 371 ATTAGTGCTATTGGAGTTTACTTATGCTCCA-----GTAAGCACATCCCC--TTGACT 321
 Qy 161 GluLeuGlnInserArgLysMetPheAlaSerMetTyrAlaLeuLysThrGlnGlyGly 180
 Db 320 CTGCAACACCAACCAAGAG-----CTAGAAGCTAAGGGTGT 282

RESULT 3
 BM132093 477 bp mRNA linear EST 27-NOV-2001
 LOCUS TSESTzyb0607.y1 TGRH Tachyzoite Subtracted cDNA Library Toxoplasma
 DEFINITION gondii cDNA clone TSESTzyb0607.y1 5', mRNA sequence.
 VERSION BM132093
 KEYWORDS EST.
 SOURCE Toxoplasma gondii
 ORGANISM Toxoplasma gondii
 Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
 Sarcocystidae; Toxoplasma.
 1 (bases 1 to 477)
 Tang, K., Cole, R., Fogarty, S., Sibley, L.D., Ajioke, J.A., White, M.,
 Clifton, S., Pape, D., Martin, J., Wylie, T., Dante, M., Marra, M.,
 Hillier, L., Kucaba, T., Theising, B., Bowers, Y., Gibbons, M.,
 Ritter, E., Bennett, J., Franklin, C., Tsagaris, V., Ronko, I.,
 Kennedy, S., Maguire, L., Waterston, R. and Wilson, R.
 Toxoplasma EST Project
 Unpublished (2001)
 Contact: Clifton, S.
 Toxoplasma EST Project
 Washington University School of Medicine
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: toxowatson.wustl.edu
 Contact David Sibley (toxowest@borcim.wustl.edu) for further
 information relating to organism, libraries, or clone availability.
 Seq primer: -40RP from Gibco
 High quality sequence stop: 420.
 Location/Qualifiers

FEATURES

source

1..477
 /organism="Toxoplasma gondii"
 /mol_type="mRNA"
 /strain="RH (Type 1)"
 /db_xref="taxon:5811"
 /clone="TSESTzyb0607.y1"
 /dev_stage="Tachyzoite"
 /lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
 /clone_lib="TGRH Tachyzoite Subtracted cDNA Library"
 /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
 XhoI; Toxoplasma RH strain tachyzoites were grown in human
 foreskin fibroblast cultures in vitro. The library was
 originally constructed by K.L.Wan, Cambridge University.
 cDNAs were synthesized from polyA RNAs by oligo d(T)
 priming and directionally cloned into the EcoRI to XhoI
 sites of the lambda ZapII vector using the Zap-cDNA
 synthesis kit (Stratagene). The primary cDNA library was
 mass excised as phagemid using ExAssist helper phage
 (Stratagene). Phagemid DNA was extracted by a pool of
 phenol-chloroform method, and hybridized against a pool of
 over-represented ESTs (N=12, from 5596 previous reads).
 The subtracted library was electroporated into DH10B
 (GeneHog, Invitrogen, Inc). WARNING: the library contains

ORIGIN

Alignment Scores:

Pred. No.:	Length:
87.00	477
Score:	Matches: 30
Percent Similarity: 43.62%	Conservative: 11
Best Local Similarity: 31.91%	Mismatches: 33
Query Match: 8.78%	Indels: 20
DB:	Gaps: 4

US-09-613-486-15 (1-198) x BM132093 (1-477)

Qy 88 GluGlyAspLeuLysLys--ValLeuAspGlyCysAlaProLeuThrArgPheThrAsn 106
 Db 140 GAGGGAGAGTTGGGAGAGATAGTCTCTTCAGCTGTGCCCTTGAGGTACACGAC 199
 Qy 107 LysLeuAArgThr-----PheGlyAArgThr 114
 Db 200 GAAAGCAGAGAGGCGGATACACATTTCTCTTCAGTCTTCATGCTTTCAGATTC 259
 Qy 115 PheThrGluAlaTyrValAspPheCysIle-----AlaTyrLysHis 128
 Db 260 TTCTCCAGACACCATTCCTTATGTATTCACAGTTGCTGGGGCTTATGAAAG 319
 Qy 129 LysLeuProGlnLeuAsnAlaAla--AlaGluLeuGlyIleProAlaGluAspSerTyr 147
 Db 320 AAACGTCCCCGGTGCATGCTGCTTTTCGCAACAAATCTGTGACGCTGAGGATCACAGA 379
 Qy 148 LeuAlaAlaAspPheLeuGlyThrCysProLysLeuSerGlu 161
 Db 380 CAGAGTACATGCACATCGCCGCAATATGTCGGGACATCACTGA 421

RESULT 4
 CF267035/c 539 bp mRNA linear EST 13-AUG-2003
 LOCUS TSESTzyj21c02.y1 TGNAS Tachyzoite cDNA Library Toxoplasma gondii
 DEFINITION cDNA clone TSESTzyj21c02.y1 5', mRNA sequence.
 VERSION CF267035
 KEYWORDS EST.
 SOURCE Toxoplasma gondii
 ORGANISM Toxoplasma gondii
 Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
 Sarcocystidae; Toxoplasma.
 1 (bases 1 to 539)
 Tang, K., Cole, R., Fogarty, S., Sibley, L.D., Ajioke, J.A., White, M.,
 Clifton, S., Pape, D., Martin, J., Wylie, T., Dante, M., Marra, K.,
 Hillier, L., Kucaba, T., Theising, B., Bowers, Y., Gibbons, M.,
 Ritter, E., Bennett, J., Franklin, C., Tsagaris, V., Ronko, I.,
 Kennedy, S., Maguire, L., Waterston, R. and Wilson, R.
 Toxoplasma EST Project
 Unpublished (2001)
 Contact: Clifton, S.
 Toxoplasma EST Project
 Washington University School of Medicine
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: toxowatson.wustl.edu
 Contact David Sibley (toxowest@borcim.wustl.edu) for further
 information relating to organism, libraries, or clone availability.
 Seq primer: -40RP from Gibco.
 Location/Qualifiers

FEATURES

source

1..539
 /organism="Toxoplasma gondii"
 /mol_type="mRNA"
 /strain="Tachyzoite"
 /db_xref="taxon:5811"
 /clone="TSESTzyj21c02.y1"
 /dev_stage="Tachyzoite"
 a small percentage of cDNAs derived from the human host
 cells. Library Source: David Sibley, Washington
 University."

/lab host="Electroten Blue cells (Stratagene)"
 /clone lib="mgms Taclyoite cDNA library"
 /note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
 XhoI; The cDNA library was constructed by Keliang Tang,
 and Robert Cole at Washington University. cDNA was
 synthesized from Poly(A)+ mRNA using an oligo-d(T) primer
 containing a XhoI site. Following second strand synthesis,
 EcoRI adapters were ligated to the cDNA, and products were
 size-selected on sephacryl S500. The cDNA were
 directionally cloned into the EcoRI/XhoI prepared
 pBluescript II SK+ vector, and electroporated into
 Electroten Blue cells (Stratagene). The library may
 contain a small percentage of host or bacterial
 contaminants."

ORIGIN

Alignment Scores:

Pred. No.: 1.49 Length: 539
 Score: 87.00 Matches: 30
 Percent Similarity: 43.62% Conservative: 11
 Best Local Similarity: 31.91% Mismatches: 33
 Query Match: 8.78% Indels: 20
 DB: 14 Gaps: 4

US-09-613-486-15 (1-198) x CF267035 (1-539)

QY 88 GUGUyAspLeuLys---ValLeuAspGlyCysAlaProLeuThrArgPheThrAsn 106
 Db 352 GAGGGAGAGTTGGGACAGATAGTCTCTTCGACGTGTCTCCCTTGAAGTACAGAAC 293
 QY 107 LysLeuArgThr-----PheGlyArgThr 114
 Db 292 GAACACAGGAGTGGGAGATACACATTCTCCCTTCAGTCTTCAGCTTTACAGATTG 233
 QY 115 PheThrGluAlaValValAspPheCysIle-----AlaTyrLysHis 128
 Db 232 TTCACACGACACACATTGCTATTGTATTCACAGTTGCTGTGGGCTTTTGAAG 173
 QY 129 LysLeuProGlnLeuAsnAlaAla---AlaGluLeuGlyLeuProAlaGluAspSerTyr 147
 Db 172 AAACGTGCCCCGTCGATGCTGCTTTTCGCGAACATACATCTGTGACGTGAGATCACA 113
 QY 148 LeuAlaAlaAspPheLeuGlyThrCysProLysLeuSerGlu 161
 Db 112 CAGACTACATCGACATCGCGCAATGTCCGGCATCATGTGA 71

RESULT 5 BQ459495 562 bp mRNA linear EST 30-MAY-2002
 LOCUS BQ459495
 DEFINITION HA08017r HA Hordeum vulgare subsp. vulgare cDNA clone HA08017

5-PRIME, mRNA sequence.

ACCESSION BQ459495.1 GI:21267266

VERSION BQ459495.1

KEYWORDS EST

ORGANISM Hordeum vulgare subsp. vulgare

SOURCE Hordeum vulgare subsp. vulgare

REFERENCE Radochuk, V., Zhang, H., Weschke, W., Potokina, E. and Wobus, U.

AUTHORS Radochuk, V., Zhang, H., Weschke, W., Potokina, E. and Wobus, U.

TITLE Barley ESTs from developing seeds

JOURNAL Unpublished (2002)

COMMENT Molecular Markers Group, Department Genbank

Institute of Plant Genetics and Crop Plant Research (IPK)

Corrensstr. 3 06466, Gatersleben, Germany

Tel: 039462-5522

Fax: 039462-5595

Email: stein@ipk-gatersleben.de

Insert Length: 562 Std Error: 0.00

Plate: 8 row: 0 column: 17

Seq primer: M13rev.

FEATURES

source

Location/Qualifiers

1..562
 /organism="Hordeum vulgare subsp. vulgare"
 /mol_type="mRNA"
 /cultiVar="barke"
 /sub_species="vulgare"
 /db_xref="taxon:112509"
 /clone="HA08017"
 /issue_type="embryosac"
 /dev_stage="0-7 DAP (days after pollination)"
 /lab_host="XU10-Gold"
 /clone_lib="HA"
 /note="Vector: pBluescript SK+; Site 1: EcoRI (5'-end of cDNA); Site 2: XhoI (3'-end of cDNA); 0-7 DAP (days after pollination) Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/White selection for recombinants is not 100% reliable. Average insert size is 1 kb"

ORIGIN

Alignment Scores:

Pred. No.: 2.49 Length: 562
 Score: 85.50 Matches: 40
 Percent Similarity: 40.00% Conservative: 38
 Best Local Similarity: 20.51% Mismatches: 84
 Query Match: 8.63% Indels: 33
 DB: 13 Gaps: 8

US-09-613-486-15 (1-198) x BQ459495 (1-562)

QY 22 GlyValAspLysLeuLeuSerAlaGluValGluLysMetLeuValGlnLysGlyAla 41
 Db 4 GGCATGAGAGAGAGCCATGATGCTACACGAGGTGCTT----- 45
 QY 42 ProAlaGluGlyIleGluValValPheGlyLeuLeuLeuTyrAlaLeuAlaAlaArgThr 61
 Db 46 ---GGAGAGGATCCAGATGATGATTTGGGCTGTACTACATCATGCTGATCG 102
 QY 62 ThrSerProLysValGlnArgAlaAspSerAspValIlePheSerAspPheGlyGlu 81
 Db 103 CTGGAGAGGACGGAGGTGATGATGATCGATGATGATGATGATGATGATGATGATGATG 162
 QY 82 -----ArgAsnValValValThrGluGlyAspLeuLysValValLeuAspGly 97
 Db 163 CATGACCTCCACAGAGATGCTGTGAATCTGTACTTTTATGATGATGATGATGCT 222
 QY 98 CysAlaProLeuThrArgPheThrAsnLysLeuArgThrPheGlyArgThrPheThrGlu 117
 Db 223 TACTGCGGTGTGAAGAAGTTCCAGATGCAATTGAGTGTGTGGCAATGCGTGAAG 282
 QY 118 AlaTyrValAspPheCysIleAlaTyr-----LysHisLys 129
 Db 283 AGTCGACACACAGATGCTCTCATACATATGCTGATTAATGCTGCGGAGAACAA 342
 QY 130 LeuProGlnLeuAsnAlaAlaGluLeu-----GlyTle--- 141
 Db 343 CTT-----GTTGTGAGAGCAGAAATTTGTCACAGATGAGGAGCGGTGATTAAT 396
 QY 142 ProAlaGluAspSerTyrLeuAlaAlaAspPheLeuGlyThrCysProLysLeuSerGlu 161
 Db 397 CCTGACGAATACCTTAATGCTTA-----CTCATGAGTCTGCTTCAAGGTTATAT 450
 QY 162 LeuGlnGlnSerArgLysMetPheAlaSerMetTyr-----AlaLeuLysThrGluGly 179
 Db 451 GTGATGACGCTGTGTGTTACTTACAGCAAGATGTTGATGTGCTCAGGCCAATGCC 510
 QY 180 GlyValValAsnThrProValSerAsnLeuArgGlnLeuGlyArg 194
 Db 511 AATGCCCTTAAACAAGATTAAGTGGCTGTGGAAGGTTGATAGG 555

RESULT 6
 BU007723 666 bp mRNA linear EST 22-AUG-2002
 LOCUS OGH5E15.V5.ab1 OG EFGHJ lettuce serritola Lactuca sativa cDNA clone
 DEFINITION OGH5E15, mRNA sequence.
 BU007723
 ACCESSION BU007723.1 GI:22442118
 VERSION
 KEYWORDS EST.
 ORGANISM Lactuca sativa
 SOURCE Lactuca sativa
 Lactuca sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
 Cichorieae; Lactuca.
 1 (bases 1 to 666)
 Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Riesberg, L.,
 Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, U., Y.,
 Ellison, P., Kolkmann, J., Slabaugh, M.S., Livingston, K., Zhou, Y.,
 Lai, Z., Church, S., Jackson, L. and Bradford, K.
 Lettuce and Sunflower ESTs from the Compositae Genome Project
 http://compositae.ucdavis.edu/
 Unpublished (2002)
 CONTACT: Alexander Kozik [R.W.Michelmore]
 Department of Vegetable Crops, R.W.Michelmore Lab
 University of California at Davis (UCD)
 Asmudson Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742
 Fax: 1-(530)-752-9659
 Email: akozik@ucdavis.edu [michelmore@vegmail.ucdavis.edu]
 belongs to contig OG_CA_Config3108, see http://cgpdb.ucdavis.edu/
 for details.
 Plate: OGH5 row: E column: 15.
 Location/Qualifiers
 1..666
 /organism="Lactuca sativa"
 /mol_type="mRNA"
 /cultivar="L.serritola"
 /db_xref="taxon:4236"
 /clone="OGH5E15"
 /lab_host="E.coli"
 /clone_lib="OG_EFGHJ lettuce serritola"
 /note="Vector: pBRCDNAseflab; The library was constructed
 from 10 different sources of RNA from a single genotype.
 Separate cDNAs were generated using primers that
 incorporated unique 5' and 3' tags to distinguish each
 source of RNA. cDNAs were then pooled, size-fractionated,
 directionally cloned into a custom medium-copy vector and
 transformations made with four size classes to minimize
 size bias. Details of each source of RNA and library
 construction can be obtained at http://cgpdb.ucdavis.edu/
 TAG_TISSUE=chemical induction
 TAG_LIB=OG EFGHJ lettuce serritola
 TAG_SEQ=TTAGCCCGG"

ORIGIN
 Alignment Scores:
 Pred. No.: 3.25 Length: 666
 Score: 85.50 Matches: 50
 Percent Similarity: 40.78% Conservative: 23
 Best Local Similarity: 27.93% Mismatches: 47
 Query Match: 8.63% Indels: 59
 DB: 13 Gaps: 13

US-09-613-486-15 (1-198) x BU007723 (1-666)
 Oy 13 ValletThraepalaSerLeuasnGlyValAspLysLysLeuLeuSerAlaGluVal 32
 Db 104 GTTTGACCGAT---GGGTCTACCATGATGTGATCCGATGCTTGGCTTCCAG--- 237
 Oy 33 GlulysMetLeuValGlnLysGlyAlaProasnGluGlyIleGluValAlaPheGlyLeu 52
 Db 238 -----CTGGCAGCAGAGAGGGCGCTTTGGAGAAAGTGT----- 270

Oy 53 LeuLeuTyraAlaLeuAlaAlaArgThrThrSerProLys-----ValGlnArg 68
 Db 271 -----AGGAAGCTGCTCCCAAAATCTTGAGCGCATATGACGA 309
 Oy 69 AlaAsp-----SerAspValIlePheSer----- 76
 Db 310 GTTGAAAGTTGTGACGCGCTGAGAACATTTGGGGAGATGATTTGGGATTTGAATTGACGA 369
 Oy 77 -----AsnSerPheGlyGluArgAsnValValValThrGluGlyAspLeuLys 92
 Db 370 AGAGGGCAGATTATACCTTTGTTGATAG-----CCTGTGACCT--- 411
 Oy 93 LysValLeuAspLysCysAlaProLeuThr-----ArgPheThrAsnLysLeuArg 109
 Db 412 AAGGTGTGATTTCTTGTGCCACCTTGGCGAAATGTTTCAGTATGTGATCTTAAGG 471
 Oy 110 -----ThrPheGlyArg-----ThrPheThrGluAlaTyrValAspPheCysIleAlaTyr 126
 Db 472 GGAATGCAAAAGGAGAGCTCTTACACATGCAATGCAATTAATTTGATGTGGCC 531
 Oy 127 LysHisLysLeuProGlnLeuAsnAlaAlaGluLeuGlyIleProAlaGluAspSer 146
 Db 532 CAACATATTGACAGACCGCTTCTGCTGTANNAGAAAGACGTGGCAGCTGATCTTCA 591
 Oy 147 -----TyrLeuAla-----AlaAspPheLeuGlyThrCys 156
 Db 592 TCTTCATCTTGTGCTATTATGATAGGGGAATTTCTGATTTTCACTTCATATGT 648

RESULT 7
 BU433495/c 671 bp mRNA linear EST 13-MAR-2002
 LOCUS BU433495 Dictyostelium discoideum cDNA library, VF Dictyostelium
 DEFINITION dictyoidem cDNA clone d4v22f05 3', mRNA sequence.
 BU433495
 ACCESSION BU433495.1 GI:19408217
 VERSION
 KEYWORDS EST.
 SOURCE Dictyostelium discoideum
 ORGANISM Dictyostelium discoideum
 Eukaryota; Mycetozoa; Dictyostelida; Dictyostellum.
 REFERENCE 1 (bases 1 to 671)
 Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
 Full length cDNA of Dictyostelium discoideum at the vegetative
 stage
 Unpublished (2002)
 CONTACT: Tadasu Shin-i
 Center for Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tsbini@genes.nig.ac.jp.
 Location/Qualifiers
 1..671
 /organism="Dictyostelium discoideum"
 /mol_type="mRNA"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="d4v22f05"
 /sex="mat A"
 /dev_stage="Growth phase"
 /clone_lib="Dictyostelium discoideum cDNA library, VF"

ORIGIN
 Alignment Scores:
 Pred. No.: 3.29 Length: 671
 Score: 85.50 Matches: 46
 Percent Similarity: 35.45% Conservative: 21
 Best Local Similarity: 24.34% Mismatches: 85
 Query Match: 8.63% Indels: 37
 DB: 12 Gaps: 5

US-09-613-486-15 (1-198) x BU433495 (1-671)

```

QY 33 GlllyMeleuValGlnlysglyAlaProasnlyglyIleGluValValPheglyLeu 52
DB 663 GAAGAGTTAAAGTTAGTGTGTAAGACACCAATTGAGAGATTGGAGATTATTA 604
QY 53 LeuLeuTyralaLeuAlaIaArgThrThrSerProlyValGlnArgAlaSerAsp 72
DB 603 TCATCATCATGATTGAGCTCATATGAC-----CAATTAACAAGTAAAGAT 556
QY 73 ValIlePheSerAsnSerPheglyGluArgAsnValValThrGlnlyAspLeu 92
DB 555 TTGATTTGACACCAAGAGTGTATTAATGTTGTCATTTGCCAATCAATTAATTAAG 496
QY 93 lyeValLeuAspGlyCysAlaProLeuThrArgPhe----- 104
DB 495 TTGGGTCTTAAAGTATTTCACCACTTCATTAACCTTTAAACAAGGTGTAACCTTCA 436
QY 105 -----ThrAsnlyLeuArgThrPheglyArgThrPheThr 116
DB 435 GTTGCTACTGATCTCCCGCTAGTAATGATTAATGATTTGGGTGATTAACCTACT 376
QY 117 GluAlaTyralaPhePheCysIleAlaTyrllyHislyLeuProGlnLeuAsnAla 136
DB 375 GCACCTTACATGATTAACCTCTCTGCAATATCTCATTT----- 334
QY 137 AlaGlnLeuGlyIleProAlaGluAspSerTyrlleu----- 148
DB 333 ---GAAGTGTGTAACCTGTAACACCATATTAAGATTTGTCAATGCACTATTAAAC 277
QY 149 AlaAlaAspPheLeuGlyThrCysProlyLeuSerGlnLeuGlnSerArglySmet 168
DB 276 GGTGCAAGGCAATGGGTATCGATGATTAAGTGTGCTCAATTTGTTAA----- 223
QY 169 PheAlaSerMetTyrlaLeuLeuThrGlnlyGlyValAlaSerThrProAlaSerAsn 188
DB 222 TTTCGATTTCAATCGCTGCTCAAGCTTCAAGTCATCGATTCATCAATCTTCAT 163
QY 189 LeuArgGlnLeuGlyArgGluVal 197
DB 162 TTAGTTATGTTGGTACTAATCATGTC 136

RESULT 8
BZ434184/ 704 bp DNA linear GSS 13-DEC-2002
LOCUS BONFR91TF_BO_1.6_2_K3_tot Brassica oleracea genomic clone BONFR91.
DEFINITION genomic survey sequence.
ACCESSION BZ434184
VERSION BZ434184.1 GI:26684219
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 704)
AUTHORS Town,C.D., Van Aken,S., Uteback,T., Koo,H. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: BONFR91TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
FEATURES
source
1..704
location/Qualifiers
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"

```

```

ORIGIN
Alignment Scores:
Pred. No.: 3.55 Length: 704
Score: 85.50 Matches: 49
Percent Similarity: 38.16% Conservative: 30
Best Local Similarity: 23.67% Mismatches: 81
Query Match: 8.63% Indels: 47
DB: 28 Gaps: 8

US-09-613-486-15 (1-198) x BZ434184 (1-704)
QY 7 SerAsnLeuSerAsnLeuValIleThrAspAlaSerSerLeuAsnlyValAlaPlySlye 26
DB 574 TCTGACCTCAACACACAGTACATGCACTTCTCTCATCTCAAGCTTGAGAGTCA 515
QY 27 Leu-----LeuSerAlaGluValGlu 33
DB 514 TTGCTTGTATGCTTATCTCTCAAGACCAATGGGGTCTCTACACAGGGAATCAGAA 455
QY 34 lyeMetLeuValGlnlyGlyAlaProasnlyglyIleGluValValPheglyLeu 53
DB 454 AAG-----AACCCCAAGAGTCTTGCAATGTTGTTCTTCC----- 419
QY 54 LeuTyralaLeuAlaIaArgThrThrSerProlyValGlnArgAlaSerAsp--- 72
DB 418 -----ACTACTCTCCAGAGATGATGATGATCATGACAA 380
QY 73 -----ValIlePheSerAsnSerPheglyGlu---ArgAsnVal 84
DB 379 GAGAGAGATGAGATTGAAGAGCTGATTTGAACTGATTTGGGAGATTGAGAGATT 320
QY 85 ValIaIThrGlnlyAspLeuIleValLeuAspGlyCysAlaProLeuThrArgPhe 104
DB 319 GTTGTCGCCACAGCTGAACACAGATTGTGATGAC---GCTGCAAGAGAGTTGAACA 263
QY 105 ThrAsnlyLeuArgThrPheglyArgThrPheThrGlnlyAlaTyralaPheCysIle 124
DB 262 ACAAAATCTGCAAGAGCTGAGACAA-----GCTGAAACAAGTTGAAGAAGAGCT 209
QY 125 AlaTyrllyHislyLeuProGlnLeuAsnAlaIaGlnlyLeuGlyIleProAlaGlu 144
DB 208 GACACCAAGCTGAATAAAGATTAAGCTTAAGAAGCACTGAGGTGAGCTTACACCTAT 149
QY 145 AspSerTyrlleuAlaIaAspPheLeuGlyThrCysProlyLeuSerGlnLeuGln 164
DB 148 GATTAAGCTCTTTTCCCAAGAGTCTCAACCAAGCTCAAGACAGAGAGTCTCCCAAG 89
QY 165 SerArglyMetPheAlaSerMetTyrlaLeuLeuThrGlnlyGlyValAlaSerThr 184
DB 88 TTCAGAAAGATCTTAAGTATTT-----GGGTGAGGCTTCA 50
QY 185 ProValSerAsnLeuArgGln 191
DB 49 GAAATCTGAGTATGCTGAA 29

RESULT 9
AU034876 713 bp mRNA linear EST 28-APR-1999
LOCUS AU034876 Dictyostelium discoideum SL (H.Urushi-hara) Dictyostelium
DEFINITION discoideum cDNA clone SL629, mRNA sequence.
ACCESSION AU034876
VERSION AU034876.1 GI:3800300
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
REFERENCE 1 (bases 1 to 713)
AUTHORS Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,

```

Yoshino, R., Mittra, B.N., Pi, M., Sato, T., Takemoto, K., Yasukawa, H., Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.
The Dictyostelium developmental cDNA project: generation and analysis of expressed sequence tags from the first-finger stage of development

JOURNAL
MEDLINE
PUBMED
COMMENT
DNA Res. 5 (6), 335-340 (1998)
99156227
10048482
Contact: Hioko Urushihara
Institute of Biological Sciences
University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
Tel: 81-298-53-4664
Fax: 81-298-53-6614
Email: hideko@biol.tsukuba.ac.jp
PROJECT = 'Dictyostelium discoideum cDNA project in Japan'
POLYA-No.

Location/Qualifiers
1. 713
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SLE629"
/dev_stage="slug"
/clone_lib="Dictyostelium discoideum SL (H. Urushihara)"

FEATURES

source

ORIGIN

Alignment Scores:

Pred. No.: 3.62 Length: 713
Score: 85.50 Matches: 46
Percent Similarity: 35.45% Conservative: 21
Best Local Similarity: 24.34% Mismatches: 85
Query Match: 8.63% Indels: 37
DB: 9 Gaps: 5

US-09-613-486-15 (1-198) x AU034876 (1-713)

QY 33 GlnlyMetLeuValGlnlyGlyAlaProasnGlnlyIleGluValAlaPheGlyLeu 52
Db 40 GAAGAGTTAAGGTTAGTGTAGAGACCAATTGAGAGATTGGAGATTGA 99
QY 53 LeuLeuTyralaLeuAlaIleArgThrThrserProlySvalGlnaGlnAlaSerSerasp 72
Db 100 TCATCATTCATTGATTGACGTCATATGACA-----CAATTAAACAAGTGAAGAT 147
QY 73 ValIlePheSerAsnSerPheGlyGlnArgasnValAlaValThrGlnGlyAspLeuLys 92
Db 148 TTAGATTGACGACCAAGAGTGTATTATTTGTTACTTGTCCAGATCAAAATTAAAG 207
QY 93 LysValIleuAspGlyCysAlaProLeuThrArgPhe----- 104
Db 208 TTGGGTGTTAAAGGATTTCACAGTTCATTAACCTTTTAAACAAGGTAAACGTTCA 267
QY 105 -----ThrAsnLysLeuArgThrPheGlyArgThrPheThr 116
Db 268 GTTGCTACTGATTCGCCCGCTAGTATATATGATTGATGTTGGGTAATTGCTACT 327
QY 117 GluAlaTyralaAspPheCysIleAlaTyrlsHisLysLeuProGlnLeuAsnAlaIa 136
Db 328 GCAGCTTCATTCATTGATAACTCTGCAAAATACCTATTCATTAAT----- 369
QY 137 AlaGluLeuGlyIleProAlaGluAspSerTyrlleu----- 148
Db 370 ---GAAGGTGGGAACCTGTAACACCATCATTAAGATTGTTGTCATAGGCACTATTAA 426
QY 149 AlaAlaAspPheLeuGlyThrCysProLysLeuSerGlnLeuGlnGlnSerArgLysMet 168
Db 427 GGTGGAAGGAGCATGGTATCGATGATAAGTTGGTTCATTCAAAATTGGTAA----- 480
QY 169 PheAlaSerMetTyralaLeuLysThrGlnGlyGlyValIleAsnThrProValSerAsn 188
Db 481 TTGGCAGATTTCATCGCTGTCAAGTTTCAAGTATCCAGTCTATGATCCATCTCTCAT 540

QY 169 LeuArgGlnLeuGlyArgArgGluVal 197
Db 541 TTAGTTAGTTGGTACTAATCATGTC 567

RESULT 10

BUJ76355/c
BUJ76355

LOCUS
DEFINITION

BUJ76355 Dictyostelium discoideum 751 bp mRNA linear EST 08-MAR-2002
discoidium cDNA clone ddc28n02 3', mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BUJ76355.1 GI:19285738
EST.
Dictyostelium discoideum
Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.

REFERENCE

1 (bases 1 to 751)
Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
Full length cDNA of Dictyostelium discoideum at the culmination stage
Unpublished (2002)

JOURNAL

Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.

AUTHORS

TITLE

FEATURES

source

Location/Qualifiers
1. 751
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddc28n02"
/sex="mat A"
/dev_stage="Culmination stage"
/clone_lib="Dictyostelium discoideum cDNA library, CF"

ORIGIN

Alignment Scores:

Pred. No.: 3.92 Length: 751
Score: 85.50 Matches: 46
Percent Similarity: 35.45% Conservative: 21
Best Local Similarity: 24.34% Mismatches: 85
Query Match: 8.63% Indels: 37
DB: 12 Gaps: 5

US-09-613-486-15 (1-198) x BUJ76355 (1-751)

QY 33 GlnlyMetLeuValGlnlyGlyAlaProasnGlnlyIleGluValAlaPheGlyLeu 52
Db 663 GAAGAGTTAAGGTTAGTGTAGAGACCAATTGAGAGATTGGAGATTGA 604
QY 53 LeuLeuTyralaLeuAlaIleArgThrThrserProlySvalGlnaGlnAlaSerSerasp 72
Db 603 TCATCATTCATTGATTGACGTCATATGACA-----CAATTAAACAAGTGAAGAT 556
QY 73 ValIlePheSerAsnSerPheGlyGlnArgasnValAlaValThrGlnGlyAspLeuLys 92
Db 555 TTAGATTGACGACCAAGAGTGTATTATTTGTTGACATTGTCAGAAATCAAAATTAAAG 496
QY 93 LysValIleuAspGlyCysAlaProLeuThrArgPhe----- 104
Db 495 TTGGGTGTTAAAGGATTTCACAGTTCATTAACCTTTTAAACAAGGTAAACGTTCA 436
QY 105 -----ThrAsnLysLeuArgThrPheGlyArgThrPheThr 116
Db 435 GTTGCTACTGATTCGCCCGCTAGTATATATGATTGATGTTGGGTAATTGCTACT 376
QY 117 GluAlaTyralaAspPheCysIleAlaTyrlsHisLysLeuProGlnLeuAsnAlaIa 136
Db 375 GCAGCTTCATTCATTGATAACTCTGCAAAATACCTATTCATTAAT----- 334

```

QY      137  AlGluLeuGlyIleProAlaGluAspSerTyrLeu----- 148
Db      333  ---GAGGTGTGTAACCTGTACACCATCATTAATAATTGTCATAGGCACTATTAC 277
QY      149  AlaAlaaspheLeuGlyThrCysProTyrLeuSerGluLeuGlnSerArgLysMet 168
Db      276  GGTGCAAGAGCATTCGCTATCCGATCGATGAATAAGCTTGTCACCTCAATGTGTAA----- 223
QY      169  PheAlaSerMetTyrAlaLeuLysThrGluGlyValValaSerThrProValSerAsn 188
Db      222  TTTCGACATTTCATCGCTGTCAAGTTTCAAGTCATCCAGTCATGATCCAACTCTCAT 163
QY      189  LeuArgGlnLeuGlyArgArgGluVal 197
Db      162  TTAGTTATGTGTGTACTATCATGTC 136

RESULT 11
BQ916284      654 bp  mRNA  linear  EST 19-AUG-2002
LOCUS      OHB17G04.YG.ab1 QG ABCDI sunflower RHA801 Helianthus annuus cDNA
DEFINITION  OHB17G04.YG.ab1 QG ABCDI sunflower RHA801 Helianthus annuus cDNA
ACCESSION  BQ916284
VERSION    BQ916284.1  GI:22315065
KEYWORDS
SOURCE
ORGANISM   Helianthus annuus (common sunflower)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Asterales; Asteraceae; Asteroideae;
            Heliantheae; Helianthus.
REFERENCE
  1 (bases 1 to 654)
  Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
  Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J.,
  Ellison,P., Kolkmann,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
  Lai,Z., Church,S., Jackson,L. and Bradford,K.
  Lettuce and Sunflower ESTs from the Compositae Genome Project
  http://compgenomics.ucdavis.edu/
  Unpublished (2002)
  Contact: Alexander Kozik [R.W.Michelmore]
  Department of Vegetable Crops, R.W.Michelmore Lab
  University of California at Davis (UCD)
  Asmundson Hall, UCD, Davis, CA 95616, USA
  Tel: 1-(530)-742-1742
  Fax: 1-(530)-752-9659
  Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
  belongs to contig QH_CA_Contig3046, see http://cgpdb.ucdavis.edu/
  for details.
  Plate: QHB17 row: G column: 04.

FEATURES
    source
        1..654
            location/Qualifiers
                /organism="Helianthus annuus"
                /mol_type="mRNA"
                /cultiivar="RHA801"
                /db_xref="taxon:4232"
                /clone="OHB17G04"
                /lab_host="E.coli"
                /clone_1ib="QG ABCDI sunflower RHA801"
                /note="Vector: pBRCNDNA51AB: The library was constructed
                from 11 different sources of RNA from a single genotype.
                Separate cDNAs were generated using primers that
                incorporated unique 5' and 3' tags to distinguish each
                source of RNA. cDNAs were then pooled, size-fractionated,
                directionally cloned into a custom medium-copy vector and
                transformations made with four size classes to minimize
                size bias. Details of each source of RNA and library
                construction can be obtained at http://cgpdb.ucdavis.edu/
                TAG_TISSUE=shoots environmental stress
                TAG_LIB=QH ABCDI sunflower RHA801
                TAG_SEQ=TCGCAACGGG"

ORIGIN
Alignment Scores:

```

```

Pred. No.: 3.67
Score: 85.00
Percent Similarity: 39.78%
Best Local Similarity: 22.65%
Query Match: 8.58%
DB: 13 Gaps: 6

```

US-09-613-486-15 (1-198) x BQ916284 (1-654)

```

QY      11  AsnLeuValIleThrAspAlaSerSerLeuAsnGlyValaSerLysLeuSerAla 30
Db      23  GATCTTGAGTCGGAGAGTAAGCCGTTATTTCGGTTATATGTCAAA-----GGCTGCT 79
QY      31  GluValGlnLysMetLeuValGlnLysGlyAlaProAsnGlnGlyIleGluValAlaPhe 50
Db      80  TCTGTAAAGCTATGCTGACATATAGGCGCTGTAATAACGAACTATTAAGTTATATAC 139
QY      51  GlyLeuLeuLeuTyrAlaLeuAlaAlaArg-----ThrThrSer 63
Db      140  GATCTTCTTGATGATGCCGAGCCGCAATGAGGACTATGACCCGTTGAGATCAA 199
QY      64  ProlLysValGlnArgAlaAspSerAspValIlePheSerAsnSerPheGlyGlu---Arg 82
Db      200  GTGCCGATTGTTGCGCAGACGTTGCGGCGGTTTGTAGCAGCGCAGTGGCGGATGGCC 259
QY      83  AsnValValValIleThrGluGlyAspLeuLysValLeuAspGlyCysAlaProLeuThr 102
Db      260  GGATCGATGATGACCTGAGGGA-----AAAGTAAACAAAGATTGGCGAGTT----- 304
QY      103  ArgPheThrAsnLysLeuArgThrPheGlyArgThrPheThrGlnAlaTyrValaSerPhe 122
Db      305  -----AGGTTTCAGAGCAAGAAACAACTTTGTTGTGTGTTCTGATTCCT 352
QY      123  CysIleAlaTyrLysHisLysLeuProGlnLeuAsnAlaAlaAlaGluLeuGlyIlePro 142
Db      353  TTGAGACGATCAAGAAAGATTTTAAAGGTAACGGCGGACCTAGATGCTGATCCGG 412
QY      413  GTTGATGATCTTATGATATGGAATGGAAGAGTATGATATCGAAGCTTTAATTCATTCAG 472
QY      143  AlaGluAsp-----SerTyrLeuAlaAlaAspPheLeuGlyThrCysProLysLeuSer 160
Db      413  GTTGATGATCTTATGATATGGAATGGAAGAGTATGATATCGAAGCTTTAATTCATTCAG 472
QY      161  GluLeuGlnLysSerArgLysMetPheAlaSerMetTyrAlaLeuLysThrGluGlyGly 180
Db      473  AAAAGCGCAACACTCGAAGAGCATCGATCATGACAGCAGCATTTAAAGAACTGGA 532
QY      181  Val 181
Db      533  ATC 535

RESULT 12
BQ848646      610 bp  mRNA  linear  EST 14-AUG-2002
LOCUS      OGATW01.YG.ab1 QG ABCDI lettuce salinas lactuca sativa cDNA clone
DEFINITION  OGATW01.YG.ab1 QG ABCDI lettuce salinas lactuca sativa cDNA clone
ACCESSION  BQ848646
VERSION    BQ848646.1  GI:22234115
KEYWORDS
SOURCE
ORGANISM   Lactuca sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
            Cichorieae; Lactuca.
REFERENCE
  1 (bases 1 to 610)
  Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
  Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J.,
  Ellison,P., Kolkmann,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
  Lai,Z., Church,S., Jackson,L. and Bradford,K.
  Lettuce and Sunflower ESTs from the Compositae Genome Project
  http://compgenomics.ucdavis.edu/
  Unpublished (2002)
  Contact: Alexander Kozik [R.W.Michelmore]
  Department of Vegetable Crops, R.W.Michelmore Lab

```


University of California at Davis (UCD)
 Asmudson Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742
 Fax: 1-(530)-752-9659
 Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
 belongs to contig QG_Ca_Contig3108, see http://cspdb.ucdavis.edu/
 for details
 Plate: QG47 row: M column: 01.

FEATURES

source
 1. 610
 /organism="Lactuca sativa"
 /mol_type="mRNA"
 /cultiivar="Salinas"
 /db_xref="taxon:4236"
 /clone="QG47M01"
 /lab_host="E.coli"
 /clone_lib="QG_ABCDI lettuce salinas"
 /note="Vector: pBRCDNA5flab; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cspdb.ucdavis.edu/TAG_LIB-QG_ABCDI lettuce salinas
 TAG_SEQ=TCGCACGGG"

ORIGIN

Alignment Scores:

Pred. No.: 3.82 Length: 610
 Score: 84.50 Matches: 50
 Percent Similarity: 40.78% Conservative: 23
 Best Local Similarity: 27.93% Mismatches: 47
 Query Match: 8.53% Indels: 59
 DB: 13 Gaps: 13

US-09-613-486-15 (1-198) x BQ848646 (1-610)

QY 13 ValIleThrsAlaSerSerLeuAsnGlyValAspLysLeuLeuSerAlaGluVal 32
 Db 137 GTTTGACCGAT---GGGTCTTACCATGATGATGATCCAGTCTCTGCTCCAG--- 190
 QY 33 GluLysMetLeuValGlnLysGlyAlaProAsnGluGlyTLeuValValPheGlyLeu 52
 Db 191 -----CTGCACGACGAGGGGGCTTTCGGGAAGTGT--- 223
 QY 53 LeuLeuTyrrAlaLeuAlaAlaArgThrSerProLys-----ValGlnArg 68
 Db 224 -----AGGAAAGCTGCTCCAAAGTCTTGGAGCCGATTATGAGA 262
 QY 69 AlaAsp-----SerAspValIlePheSer----- 76
 Db 263 GTTGAATTGTGACGCTGGAAGAACATTTGGGGGATGTGATGGGATTGAAATTCGGGA 322
 QY 77 -----AnSerPheGlyGluArgAsnValValThrGluGlyAspLeuLys 92
 Db 323 AGAGGGCAGATTATATACCTTTGGATAG-----CCTGGTGAAT--- 364
 QY 93 LysValLeuAspGlyCysAlaProLeuThr-----ArgPheThrAsnLysLeuArg 109
 Db 365 AAGGTGTGATTTCTTGGTCCCACTTGGGAATGTTCAAGATATGAGTACTTTAAG 424
 QY 110 -----ThrPheGlyArg-----ThrPheThrGluAlaTyrrValAspPheCysIleAlaTyrr 126
 Db 425 GGAATGACAAAGAGAGAGCTTTTACACATGCAATGCAATTAATTGATGTGCTT 484
 QY 127 LysHisLysLeuProGlnLeuAsnAlaAlaAlaGluLeuGlyLeuProAlaGluAspSer 146
 Db 485 CACATATTCAGAACACGCTTTCTGCTGCCAAAGAGAAAGAGCTGTGACAGCTTATCTTCA 544

QY 147 -----TyrLeuAla-----AlaAspPheLeuGlyThrCys 156
 Db 545 TCTTCACTTTGTGTCATTTTATGATAGAGGAGATTTTCTGATTTTCACTTCATATGT 601

RESULT 13

BQ984758

LOCUS

DEFINITION

OG55A05.YG.ab1 QG_EFGHU lettuce serritola Lactuca sativa cDNA clone

OG55A05 mRNA sequence.

ACCESSION

VERSION

KEYWORDS

ORGANISM

SOURCE

Lactuca sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Asterids; campanulids; Asteriales; Asteraceae; Cichorioideae;

Cichorieae; Lactuca.

1 (bases 1 to 611)

Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,

Liu,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J.,

Ellison,P., Kolkman,J., Stabaugh,M.S., Livingston,K., Zhou,Y.,

Lai,Z., Church,S., Jackson,L. and Bradford,K.

Lettuce and Sunflower ESTs from the Compositeae genome project

http://compenomics.ucdavis.edu/

Unpublished (2002)

Contact: Alexander Kozik [R.W.Michelmore]

Department of Vegetable Crops, R.W.Michelmore Lab

University of California at Davis (UCD)

Asmudson Hall, UCD, Davis, CA 95616, USA

Tel: 1-(530)-742-1742

Fax: 1-(530)-752-9659

Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]

belongs to contig QG_Ca_Contig3108, see http://cspdb.ucdavis.edu/

for details.

Plate: QG55 row: A column: 05.

Location/Qualifiers

1. 611

/organism="Lactuca sativa"

/mol_type="mRNA"

/cultiivar="L. serritola"

/db_xref="taxon:4236"

/clone="OG55A05"

/lab_host="E.coli"

/clone_lib="QG_EFGHU lettuce serritola"

/note="Vector: pBRCDNA5flab; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cspdb.ucdavis.edu/TAG_LIB-QG_EFGHU lettuce serritola
 TAG_SEQ=TGTAGCCGGG"

ORIGIN

Alignment Scores:

Pred. No.: 3.83 Length: 611
 Score: 84.50 Matches: 50
 Percent Similarity: 40.78% Conservative: 23
 Best Local Similarity: 27.93% Mismatches: 47
 Query Match: 8.53% Indels: 59
 DB: 13 Gaps: 13

US-09-613-486-15 (1-198) x BQ984758 (1-611)

QY 13 ValIleThrsAlaSerSerLeuAsnGlyValAspLysLeuLeuSerAlaGluVal 32
 Db 137 GTTTGACCGAT---GGGTCTTACCATGATGATGATCCAGTCTCTGCTCCAG--- 190

QY 33 GtUleuMetLeuValGlnUlysglyAlaProAsnGlyIleGluValValPheGlyLeu 52
 Db 191 -----CTGGCAGACGACGAGCGCTTTCGGAGAGTGCTT----- 223
 QY 53 LeuLeuTyralaLeuAlaIaArgThrThrserProlys-----ValGlnArg 68
 Db 224 -----AGGAAAGCTGCTCCAAAGATCTTGAGAGCCGATTAAGAGA 262
 QY 69 AlaAsp-----SerAspValIlePheSer----- 76
 Db 263 GTTGAAGTGTGACGCGCTGAGAAACATTTGGGGATGTGATGGAGATTGAATTCGAGA 322
 QY 77 -----AnserPheGlyGluArgAsnValValThrGluGlyAspLeuLys 92
 Db 323 AGAGGCGAGATTATATACCTTGTGGATAC-----CCTGGTGACCTT--- 364
 QY 93 LysValLeuAspGlyCysAlaProLeuThr-----ArgPheThrAsnLysLeuArg 109
 Db 365 AAGGTGTGATTTCTTGTGCCACTTGGCGAAATGTTCAAGATGAGTACTTTAAG 424
 QY 110 -----ThrPheGlyArg-----ThrPheThrGluAlaIaTyralaAspPheCysIleAlaTy 126
 Db 425 GGAATGCAAAAGGAGAGAGCTTCTTACCAATGCAATTAATTTGATGTGGCT 484
 QY 127 LysHisLysLeuProGlnLeuAsnAlaAlaGluLeuGlyIleProAlaGluAspSer 146
 Db 485 CAACATATTCAGAACCGACTTCTGCTGCTTAAGAAAGAGAGCTGGAGCTTATCTCA 544
 QY 147 -----TyrlaAla-----AlaAspPheLeuGlyThrCys 156
 Db 545 TCTTCATCTTGTGCTATTTAATGATAGGAGAAATTTCTGATTTTCACTTCATATGT 601

RESULT 14
 BG526441 632 bp mRNA linear EST 16-NOV-2001
 LOCUS 61-95 Stevia field grown leaf cDNA Stevia rebaudiana cDNA 5', mRNA
 DEFINITION
 ACCESSION BG526441
 VERSION BG526441.1 GI:16949914
 KEYWORDS EST.
 SOURCE
 ORGANISM Stevia rebaudiana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; campanulids; Asterales; Asteraceae; Asteroidae;
 Eupatoriaceae; Stevia.
 1 (bases 1 to 632)
 Brande, J.E., Richman, A., Swanson, A.K. and Chapman, B.P.
 Leaf ESTs from Stevia rebaudiana: A resource for gene discovery in
 diterpene synthesis
 Unpublished (2001)
 JOURNAL
 COMMENT Contact: Jim Brande
 Genomics and Biotechnology
 Agriculture and Agri-Food Canada - SCPPRC
 1391 Sandford St., London, Ontario, CANADA, N5V 4T3
 Tel: 519 457 1470
 Fax: 519 457 3997
 Email: brandeje@em.agr.ca
 Seg primer: T3 promoter primer.
 Location/Qualifiers
 1..632
 /organism="Stevia rebaudiana"
 /mol_type="mRNA"
 /strain="751/1501"
 /cultivar="Landrace"
 /dd_xref="taxon:55670"
 /tissue_type="leaf"
 /dev_stage="field grown, mid-size"
 /lab_host="E. coli strain XL0LR"
 /clone_id="Stevia field grown leaf cDNA"
 /note="Vector: pBK-CMV; Site 1: EcoRI; Site 2: XhoI; This
 cDNA library was constructed from polyA+ enriched mRNA
 from field grown leaves. Mid-size actively growing leaves

ORIGIN
 Alignment Scores:
 Pred. No.: 4.04 Length: 632
 Score: 84.50 Matches: 48
 Percent Similarity: 40.57% Conservative: 23
 Best Local Similarity: 27.43% Mismatches: 47
 Query Match: 8.53% Indels: 57
 Db: 12 Gaps: 12

US-09-613-486-15 (1-198) x BG526441 (1-632)

QY 2 GtUleuMetSerAspSerAsnLeuSerAsnLeu-----ValIle 14
 Db 62 GAGTCAATGTCTTAATGAGATCTGCGGGCTTCCCTGTGTGACCTTGTGCGGTGTG 121
 QY 15 ThrAspAlaSerSerLeuAsnGlyValAspLysLysLeuSerAlaGluValGly 34
 Db 122 ACCGAT---GGCTCATACCATGATGTTACTCGAGTCACTCGGCTCCAG--- 169
 QY 35 MetLeuValGlnUlysglyAlaProAsnGlyIleGluValValPheGlyLeuLeu 54
 Db 170 ---CTGGCGGCTCGAGAGCTTCCGTGAAGAGATT--- 202
 QY 55 TyrlaLeuAlaAlaArgThrThrserProlys-----ValGlnArgAlaAsp 70
 Db 203 -----AGAAAGCGGGTCCAAAGATTTTGAACCCATATAGAGTTGAA 247
 QY 71 -----SerAspValIlePheSer----- 76
 Db 248 GTTGTGACACCTGAGAAACATTTGGGTGATGTATGTGATTAATTCAGAAAGAGA 307
 QY 77 -----AnserPheGlyGluArgAsnValValThrGluGlyAspLeuLysVal 94
 Db 308 CAATTAATAGCTTGTGGAGCAA-----CCTGGTGCTC---AAGGTG 349
 QY 95 LeuAspGlyCysAlaProLeuThr-----ArgPheThrAsnLysLeuArg----- 109
 Db 350 GTGGATTACTTGTGCCACTGACAGAAATGTTCAATATGAGAGCTTAAGAGAAATG 409
 QY 110 ThrPheGlyArg-----ThrPheThrGluAlaTyralaAspPheCysIleAlaTyrlaHis 128
 Db 410 ACMAAAGGTGCGACATCTTACCTATGCACTAGCAATGCAAAATTCATGATGTGACTTCAAC 469
 QY 129 LysLeuProGlnLeuAsnAlaAlaGluLeuGlyIleProAla 143
 Db 470 ATTCAAGACCAACTTCTGCTGCCAAAGAGAGAGTACAGCT 514

RESULT 15
 BG523153 648 bp mRNA linear EST 16-NOV-2001
 LOCUS 29-48 Stevia field grown leaf cDNA Stevia rebaudiana cDNA 5', mRNA
 DEFINITION
 ACCESSION BG523153
 VERSION BG523153.1 GI:16946562
 KEYWORDS EST.
 SOURCE
 ORGANISM Stevia rebaudiana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; campanulids; Asterales; Asteraceae; Asteroidae;

REFERENCE
1 (bases 1 to 648)
Eupatoriaceae; Stevia.
Brandle, J.E., Richman, A., Swanson, A.K. and Chapman, B.P.
Leaf ESTs from Stevia rebaudiana: A resource for gene discovery in
diterpene synthesis
Unpublished (2001)
JOURNAL
Contact: Jim Brandle
Genomics and Biotechnology
Agriculture and Agri-Food Canada - SCFRC
1391 Sandford St., London, Ontario, CANADA, N5V 4T3
Tel: 519 457 1470
Fax: 519 457 3997
Email: brandleje@em.agr.ca
COMMENT
Seq primer: T3 promoter primer.
Location/Qualifiers

FEATURES

1..648
/organism="Stevia rebaudiana"
/mol_type="mRNA"
/strain="751/1501"
/cultivar="Landrace"
/db_xref="taxon:55670"
/tissue_type="leaf"
/dev_stage="field grown, mid-size"
/lab_host="E. coli strain XL0LR"
/clone_lib="Stevia field grown leaf cDNA"
/note="Vector: PBK-CMV; Site 1: EcoRI; Site 2: XhoI. This
cDNA library was constructed from polyA+ enriched mRNA
from field grown leaves. Mid-size actively growing leaves
were collected and pooled from several plants and frozen
immediately after harvesting in liquid nitrogen. The cDNA
was prepared using an XhoI-poly(dT) linker-primer. An
EcoRI adapter was ligated to the blunt end cDNA and the
products were digested with EcoRI and XhoI enabling
directional cloning into the lambda ZAP Express vector.
The library was amplified using the host strain XL1-Blue
MRF. Mass excision of the library was performed to
obtain PBK-CMV phagemid clones in the host strain XL0LR.
Single pass DNA sequencing was performed using the T3
promoter primer: 5' ATTAACCTCCTCAAGGA 3'. This library
was constructed by Alex Richman."

ORIGIN

Alignment Scores:

Pred. No.: 4.2 Length: 648
Score: 84.50 Matches: 48
Percent Similarity: 40.57% Conservative: 23
Best Local Similarity: 27.43% Mismatches: 47
Query Match: 8.53% Indels: 57
DB: 12 Gaps: 12

US-09-613-486-15 (1-198) x BG523153 (1-648)

QY 2 GluLeuWetSerAspSerAsnLeuSerAsnLeu-----ValIle 14
DB 80 GAGTCAATGCTTAATGAGTACTCGCGGTTTCCTGTTGACCTTCGTCGGTGTG 139
QY 15 ThrAspIaSerSerLeuAsnGlyValAspIysIysLeuSerAlaGluValGluLys 34
DB 140 ACCGAT---GGCTCTTACCATGATGATTTGACTCGAGTACTCGCGTCCAG----- 167
QY 35 MetLeuValGluIysGlyValAspAsnGluGlyIleGluValValPheGlyLeuLeuLeu 54
DB 188 ---CTGGCGGCTCGAGGAGCTTTCGTTGAGAGATT----- 220
QY 55 TyrAlaLeuAlaAlaArgThrThSerProLys-----ValGluArgAlaAsp 70
DB 221 -----AGAAAGCGGGTCCCAAGATTTTGCAGAACCCATTATGAGAGTTGAA 265
QY 71 -----SerAspValIlePheSer----- 76
DB 266 GTTGTGACACCTGAGAGACATTGGGTGATGTATTTGATTTGAATTGATTCAGAGAGA 325
QY 77 -----AsnSerPheGlyGluArgAsnValValValThrGluGlyAspLeuLysLysVal 94

DB 326 CAAATTAAATAGCTTTGGTGAACAA-----CTTGAGGCTC---AAGGTG 367
QY 95 LeuAspGlyCysAlaProLeuThr-----ArgPheThrAsnLysLeuArg----- 109
DB 368 GTGGATTCACCTGTGGCCACTAGAGAAATGTTTCATATGTGAGTAAAGAGATG 427
QY 110 ThrPheGlyArg---ThrPheThrGluAlaIleTyrValAspPheCysIleAlaIleTyrLysHis 128
DB 428 ACAAAAGGTGAGAGATCTTACACTATGCACTAGCAAAATTCATGTGTGATCCTCAAC 487
QY 129 LysLeuProGluLeuAsnAlaAlaAlaGluLeuGlyIleProAla 143
DB 488 ATTGAGAACCACTTCTGCTGCCAAGAAAGAGACAGTTACAGCT 532

Search completed: April 7, 2004, 03:41:23
Job time : 2366 secs